Structural studies of the recognition of bacterial lipopolysaccharides by human surfactant protein-D

Caroline Smallcombe

Thesis for submission towards the degree of Doctor of Philosophy

December 2016

Keele University

Keele University

SUBMISSION OF THESIS FOR A RESEARCH DEGREE

Part I. DECLARATION by the candidate for a research degree. To be bound in the thesis

Degree for which thesis being submitted: PhD

Title of thesis: Structural studies of the recognition of bacterial lipopolysaccharides by human surfactant protein-D

This thesis contains confidential information and is subject to the protocol set down for the submission and examination of such a thesis.

YES/~~NO~~ [please delete as appropriate; if YES the box in Part II should be completed]

Date of submission: 12th April 2016 Original registration date: 27th September 2010

(Date of submission must comply with Regulation 2D)

Name of candidate: Caroline Smallcombe (95007629/2)

Research Institute: FNS

Name of Lead Supervisor Prof. T J Greenhough

I certify that:

(a) The thesis being submitted for examination is my own account of my own research
(b) My research has been conducted ethically. Where relevant a letter from the approving body confirming that ethical approval has been given has been bound in the thesis as an Annex
(c) The data and results presented are the genuine data and results actually obtained by me during the c onduct of the research
(d) Where I have drawn on the work, ideas and results of others this has been appropriately acknowledged in the thesis
(e) Where any collaboration has taken place with one or more other researchers, I have included within an 'Acknowledgments' section in the thesis a clear statement of their contributions, in line with the relevant statement in the Code of Practice (see Note overleaf).
(f) The greater portion of the work described in the thesis has been undertaken subsequent to my registration for the higher degree for which I am submitting for examination
(g) Where part of the work described in the thesis has previously been incorporated in another thesis submitted by me for a higher degree (if any), this has been identified and acknowledged in the thesis
(h) The thesis submitted is within the required word limit as specified in the Regulations

Total words in submitted thesis (including text and footnotes, but excluding references and appendices) …38,378……

Signature of candidate ……C. C. Smallcombe…… Date …10/11/2016…

Acknowledgements

I would like to express my total gratitude to Trevor Greenhough and the Structural Biology group at Keele for enabling me to complete my thesis. Trevor's patience and quiet understanding will never be forgotten, and I hope I absorbed even a fraction of his scientific knowledge. In addition, and in no particular order, my thanks to Jenny Moran, Jenny Paterson, Ian Burns, Ruben Da Silva, Rob Williams, Jamie Littlejohn and Darius McCleary - all of whom contributed to making these past few years not only possible, but enjoyable.

To my office-mates Zainab, Rawaa and Wafaa, whose ceaseless kindness and support taught me nothing of science, but much about life - I offer my love and total thanks; and to my friend Leanne Patrick, who understands. I would also like to thank my father for his unwavering financial support, and my brother for his constant light-heartedness; without both of whom I would have forgotten there was, in fact, a light at the end of the tunnel.

Last but not least, my eternal appreciation and gratefulness to Mike Nicosia. Although I didn't always show it, I was kept (somewhat) sane - I owe you.

Abstract

SP-D is a large hydrophilic protein, consisting of four collagenous trimers, which is secreted by alveolar type II and non-ciliated bronchiolar epithelial cells and participates in calcium-dependant agglutination of inhaled microorganisms. Known high-resolution crystal structures of mono- and disaccharide bound recombinant human SP-D include a heptose disaccharide that mimics the inner core fragment of bacterial lipopolysaccharide. This thesis focuses on the structural characterisation of SP-D binding to lipopolysaccharides from Gram-negative bacteria. Intact LPS and several hydrolysed smooth and rough mutants were soaked into native crystals of a recombinant head and neck fragment of SP-D. Those soaked with hydrolysed *Escherichia coli* J5 and *Salmonella minnesota* R7 LPS showed electron density corresponding to ligand in the ligand-binding site. All crystals processed conformed to spacegroup $P2_1$, all with unit cells close to a= 55, b= 108, c= 55 Å, β= 91°. The maximum resolution diffraction observed was 1.63 Å.

The R7-soaked structure was refined at 1.77 Å, with a final R-factor of 18.71% and R-free of 22.05%. The structure reveals a well-defined trisaccharide unit of two heptose saccharides and a single Kdo residue in protein chains B and C. The Kdo is present as a five-membered, anhydro residue known to form during mild acid hydrolysis. The third, outermost heptose of the R7 inner core is not visible in the electron density.

The refined structure demonstrates binding of LPS via the O6 and O7 hydroxyl groups of the glycerol sidechain of HepI, the innermost heptose, demonstrating for the first time structurally not only the binding of a physiologically relevant bacterially derived ligand but also the recognition of a non-terminal monosaccharides by SP-D. No direct interaction is observed between the second heptose and the protein, but two hydrogen bonds are seen between the anhydro-Kdo and the amino acids Arg343 and Asp325 which flank the SP-D ligand binding pocket.

Contents

List of figures

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

List of tables

Chapter 1

Chapter 2

Chapter 4

Chapter 5

Chapter 1: Introduction to Collectins: pathogen recognition molecules

1.1 The immune system

The mammalian immune system exists as two distinct mechanisms - the evolutionarily ancient *innate* immune system and an *adaptive* or *acquired* immune system. The former acts as the first line of defence for all species, characterised by rapid target recognition of general non-self-antigenic moieties, while the latter arose during early vertebrate evolution to detect specific microbial antigens. The acquired immune system requires an initial encounter with the invading pathogen to provide protection against re-infection and has evolved to provide a more versatile means of defence. Activation is slow however, and begins when a pathogen is ingested by dendritic cells in the infected area. Precursor cells such as macrophages then migrate from the bone marrow and 'survey' the local environment. Eventually dendritic cells migrate through the lymphatic system to lymph nodes where they interact with lymphocytes. If the dendritic cells fail to become activated, the immune system is said to be tolerant of the antigen. The function of dendritic cells is not to directly neutralise the invading pathogen, but to present the antigen to peripheral lymphoid organs where it becomes an *antigen-presenting cell*. All adaptive immune responses are mediated by lymphocytes. Lymphocytes are a class of white blood cells that bear cell-surface receptors for an antigen. There are two main classes of lymphocyte: B lymphocytes (B cells) which mediate humoral responses in which antibodies produced by B cells cause the destruction of extracellular microorganisms and prevent the spread of intracellular infections; and T lymphocytes (T cells) which evoke cell-mediated immunity, whereby infected cells are destroyed by cytotoxic T cells, or intracellular pathogens are destroyed by macrophages activated by T-helper, T_H1, cells.

Lymphocytes that have been activated by an antigen give rise to clones of antigen-specific cells that mediate adaptive immunity. The result is production of antibodies of different specificities that serve as antigen receptors. The process is therefore not instantaneous. After a naive

lymphocyte has been activated, it takes 4 to 5 days before clonal expansion is complete and the lymphocytes have differentiated. When a recirculating lymphocyte encounters its specific foreign antigen it proliferates and its progeny then differentiate into effector cells that can eliminate the infectious agent. A subset of these proliferating lymphocytes differentiate into memory cells, ready to respond rapidly to the same pathogen if it is encountered again (Janeway, 2001).

1.1.1 Innate immunity

Given the time taken for the adaptive immune system to eliminate infectious agents, the importance of a rapid response is clear. Most infectious agents induce immune responses by activating innate immunity, as microorganisms such as bacteria that penetrate the epithelial surfaces of the body are met immediately by cells and molecules that elicit this response. For example, phagocytic macrophages have surface receptors able to recognise and bind common components of many bacterial surfaces. Binding of these moieties triggers the macrophage to engulf the bacterium and induce secretion of biologically active molecules such as cytokines and chemokines that affect the behaviour of other cells that bear receptors for them. Chemokines attract cells such as neutrophils and monocytes from the bloodstream and initiate inflammation, which may trigger a complement cascade on the bacterial cell surface. Examples of innate immunity in the form of chemical and physical barriers are vast: the skin serves as a primary barrier to microbial entry; epithelial cells produce secretions which contain antimicrobial peptides and enzymes that inhibit pathogen growth; the mucus lining of the throat and lungs acts as a physical barrier that prevents adherence of pathogens and cilia facilitate clearance; orifices such as the eyes and mouth contain enzymes such as lysozyme and continuously wash away microbes; while highly acidic secretions in the stomach ensure few organisms survive if ingested.

Despite these barriers, pathogens may still enter the body and once inside, other mechanisms induce further innate immune responses. Pattern recognition receptors (PRRs) such as Toll-like

receptors (TLRs), Nod-like proteins, and collectins act as sensors of intracellularly encountered microbial motifs (Ozinsky *et al.,* 2000, Fritz *et al.,* 2006; Lu, 2002). Present round the body in the form of soluble proteins and cell surface receptors, PRRs induce various signalling pathways including opsonisation of the microbe, recruitment of phagocytic and natural killer cells, and upregulation of the complement cascade (Medzhitov & Janeway, 2000).

These molecules are able to distinguish microbial pathogens from healthy self-antigens through recognition of molecular arrays known as pathogen-associated molecular patterns (PAMPs), which are highly conserved among the different classes of pathogens. However, invading pathogens are not the only causative agents of cell damage; trauma and associated tissue-damage are also recognised at the cellular level by receptor-mediated detection. For instance, intracellular proteins released by dead or apoptotic cells elicit similar responses, and are referred to as damage-associated molecular patterns, or DAMPs (Bianchi, 2006).

1.1.2 Lectins

Molecules that recognise carbohydrate moiety PAMPs are termed lectins, and are classified into distinct subgroups on the basis of their amino acid sequences and biochemical properties: *P-type lectins* such as the mannose 6-phosphate receptor and the insulin-like growth factor II/mannose 6-phosphate receptor which recognise and bind phosphorylated mannose residues (Dahms, 2002); *I-type lectins*, glycan-binding proteins that belong to the immunoglobulin superfamily (Varki, 1999); *Gal-lectins* which bind galactose and N-acetylgalactosamine (Boettner *et al.,* 2002); and *C-type lectins* whose members contain carbohydrate-recognition domains (CRDs) that are homologous even though they mediate a variety of interactions by binding selectively to specific monosaccharides.

C-type lectins are further divided in to two distinct subgroups: *selectins* which bind fucosylated and sialylated glycoproteins, found on endothelial cells, leukocytes and platelets (Ley, 2003), and *collectins*.

1.1.3 The Collectin family

The collectins are calcium-dependent pattern recognition molecules found in mammals, and represent a non-clonal, innate host defence system. Various collectins have been characterised in the following major locations: lung surfactant proteins A and D (SP-A, SP-D), found on the epithelial lining of lungs to provide protection against inhaled pathogens and allergens; hydrophobic surfactant proteins B and C (SP-B, SP-C) which stabilise the respiratory surface of mammalian lungs (Parra *et al*., 2013); mannose-binding lectin (MBL) found in the serum and liver (Weis *et al*., 1998); conglutinin, collectin-43 and collectin-46, bovine serum proteins closely related to SP-D (Håkansson & Reid, 2000); and the more recently discovered collectin-liver (CL-L1), collectin-kidney (CL-K1) and collectin-placenta (CL-P1) molecules (Ohtani *et al.,* 1999; Jang *et al.,* 2008; Axelgaard *et al*., 2013). Each are involved in protein-carbohydrate interactions and have important roles in two distinct aspects of the innate immune response: initial pathogen recognition and the subsequent cellular interactions that lead to the pathogen's neutralisation and clearance. The localisation and carbohydrate binding preference of these molecules are summarised in the following table:

Table 1-1 Localisation of collectins in tissues, and relative carbohydrate specificity. Lu *et al.*, 2002.

Collectin	Localisation in tissues	Carbohydrate specificity
SP-A	Lung epithelium, prostate, thymus, intestinal mucosa, Eustachian tube, middle ear, paranasal sinuses, mesothelium, synovial fluid	ManNAc>L-fucose/maltose> glucose>mannose/galactose
SP-D	Lung epithelium, gastrointestinal epithelium, kidney, brain, testis, pancreas, salivary gland, heart, prostate, small intestine, urinary tract, placenta, uterus, stomach, mammary glands, spleen, adrenal gland, and liver	maltose>L-fucose>mannose> Glucose>>>glucosamine
MBL	Serum and liver	GlcNAc>mannose/L-fucose> ManNAc>>>maltose>glucose
Conglutinin	Bovine liver	GlcNAc>>>mannosamine> L-fucose/mannose>glucose>
CL-43	Bovine liver	mannose>>>ManNAc>>>L-fucose>GlcNAc>glucose> galactose

Table 1-2 Localisation of collectins in tissues, and relative carbohydrate specificity. Lu *et al.*, 2002.

Collectin	Localisation in tissues	Carbohydrate specificity
CL-46	Bovine thymus, liver, mammary glands and digestive system	GlcNAc>>>ManNAc /mannosamine>>maltose> glucose/L-fucose/mannose>>galactose
CL-L1	Tissues except skeletal muscle	mannose
CL-K1	Kidney, vascular endothelial cells and placenta	fucose>>mannose
CL-P1	Placenta, alveolar macrophages and tonsils	lactose> GalNAc>L-fucose>galactose

ManNAc - N-acetyl mannosamine; GlcNAc - N-acetyl glucosamine; GalNAc - N-acetyl galactosamine

Collectins target non-self pathogens such as viruses, bacteria, fungus and allergens by binding to their characteristic surface carbohydrates, and the broad range of recognised monosaccharides reflects the variety of pathogenic surfaces to which they bind. Binding is mediated by the conserved calcium-recognition domain. Collectins bind glycoconjugates and lipids moieties on a broad range of microorganisms, fungi and allergens, as well as receptors on host cells. Through these interactions, a range of effector mechanisms are induced, such as agglutination through formation of bridges between ligands on a pathogen's surface (van Rozendaal *et al.,* 2000); activation of the complement cascade (Hajela, 2002); opsonisation and activation of phagocytosis by binding of receptors on phagocyte cell surface (Beharka *et al.,* 2002); inhibition of microbial growth by increasing cell permeability of the bacterial membrane (Wu *et al.,* 2003); modulation of the inflammatory response by recruitment of chemokines, cytokines and production of reactive oxygen and nitrogen species (Tino & Wright, 1998); modulation of the adaptive immune response by inhibition of T-cell proliferation (Borron *et al.,* 1998); modulation of the allergic response by inhibition of IgE binding to allergens and subsequent suppression of histamine release (Strong *et al.,* 2002); stimulation of apoptotic cell clearance by alveolar macrophages (Schagat *et al*., 2001);

as well as various methods of inhibiting uptake of microorganisms by self-cells (Hartshorn *et al.,* 2006).

1.1.3.1 Collectin structure

Soluble collectins are large homo-oligomers that comprise a collagenous N-terminal region, a coiled-coil neck and a C-terminal calcium recognition domain (Drickamer & Taylor, 1993). Two subgroups of collectins are distinguished by the length of the collagenous domain: those with a smaller number of Gly-X-Y repeats in the collagen domain form 'bouquet'-like structures, while those with more repeats form cruciform (see Fig 1-1).

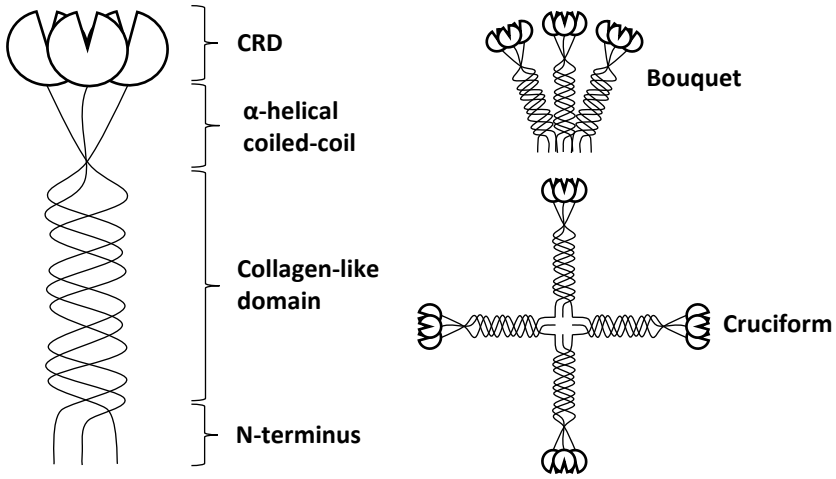


Figure 1-1 Domain organisation of collectins. The left structure shows a single trimeric building block and its constituent parts, the right represents the bouquet and cruciform oligomers. Adapted from Weis *et al.,* 1998.

Collectins contain four distinct regions of primary structure: a short amino-terminus that form inter- and intra-chain disulphide bonds that stabilise the trimeric building block (Wallis & Drickamer, 1997); a section of Gly-X-Y repeats that form a collagen domain; an α-helical coiled-coil neck of about 30 amino acids; and a carboxy-terminal carbohydrate recognition domain (CRD).

The following diagram (Fig 1-2) shows a sequence alignment, comparing the head and neck region of the four aforementioned human collectins SP-A, SP-D, MBL and CL-L1:

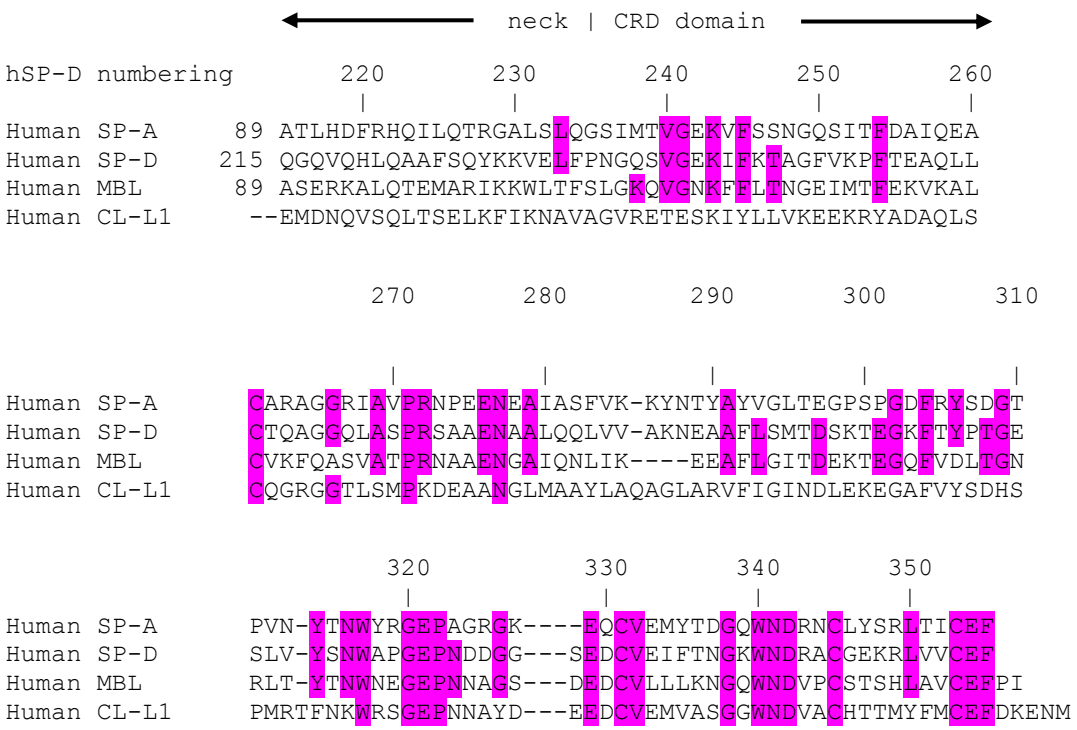


Figure 1-2 Sequence alignment diagram of the neck and head domains of some human collectins. Sequences were taken from UniProt.org and aligned using ClustalW2 (Larkin *et al.*, 2007).

1.1.3.1.1 N-terminus

The N-terminal region which stretches from the N-terminus to the first collagenous triple-helix residue, a cysteine-rich domain that stabilises trimers through disulphide bridging that links the collectin oligomers together (Holmskov *et al.,* 1995). There is no overall homology between collectins in the N-terminal region, but collectins can be divided into four subclasses: in SP-D, conglutinin and CL-43 they are similar in length at 25-28 amino acids; SP-A has a much shorter domain, one isoform with seven residues, the other consisting of ten; the third group, which includes MBL, is capable of forming larger oligomers; and the final group, so far only represented by CL-L1, has a much longer N-terminal sequence. Fig 1-3, below, demonstrates the lack of similarity between N-terminal regions of human collectins SP-A, SP-D, MBL and CL-L1:

```
Human SP-A      ------------------------EVKDVCVGSPGIPGTPGSHGLPGRDG
Human SP-D      ---------------AEMKTYSHRTMPSACTLVMCSSVESGLPGRDGRDG
Human MBL       ------------------ETVTCEDAQKTCPAVIACSSP-GINGFPGKDG
Human CL-L1     MNGFASLLRRNQFILLVLFLLQIQSLGLDIDSRPTAEVCATHTISPGPKG
```

Figure 1-3. Sequence alignment diagram of the N-terminal domains of some human collectins. The N-terminal region is highlighted in yellow, and cysteine residues are pink. Sequences were taken from UniProt.org and aligned using ClustalW2 (Larkin *et al*., 2007).

1.1.3.1.2 Collagen region

The presence of a collagen-like region is necessary for the protein's trimeric structure, which is essential for proper function. High-order organisation strengthens the carbohydrate-protein interactions - and as such assembly of collectins in to larger entities allows cross-linking of several target particles, and permits the multiple CRDs to interact simultaneously with a single microbe (Håkansson *et al.,* 2000). There are however, no experimentally determined structures of native collagens, and current knowledge is based mainly on fibre diffraction studies, model building and crystallographic studies of short compounds (Kramer *et al.,* 1998). Collagen structures can be recognised by their Gly-X-Y repeat pattern, where X and Y can be any amino acid but are typically proline, and form a left-handed helix. The three oligomers coil round one another in a right-handed superhelix and interchain hydrogen bonds between Gly amine groups and X carbonyl groups stabilise the structure (Kramer *et al.,* 1999). The helix is further stabilised by a network of water molecules that surround it. Its high tensile-strength and relative resistance to proteolysis mark the collagen region as an ideal domain cross-linker. In addition, this region has been shown to mediate binding to other macromolecules, such as the C1q receptor and macrophage scavenger receptor (Stuart *et al.,* 1997).

1.1.3.1.3 α-helical coiled-coil neck

α-helices often show a distinct heptad repeat pattern a-b-c-d-e-f-g, where a and d are hydrophobic amino acids that form the interior of the coil. α-helices typically make 8 or 9 turns in

total (Weis & Drickamer, 1994) and offer an explanation for the ability of collectins to distinguish self from non-self: a hydrophobic interface between the neck and C-terminal CRD maintains a fixed spatial relationship, such that the binding sites are roughly 45 Å apart in the trimer. Pathogenic surfaces present dense, repetitive arrays of ligands that span this distance perfectly (Weis & Drickamer, 1994). It is thought that the primary role of the neck domain in molecular assembly is to align the collagen chains and thereby facilitate subsequent folding of the collagen helix (Zhang *et al.,* 2001).

1.1.3.1.4 Carbohydrate recognition domain

The carbohydrate recognition domain (CRD) is a well conserved domain of 110-140 amino acids, characterised by a double lobe structure (see Fig 1-4, panel A).

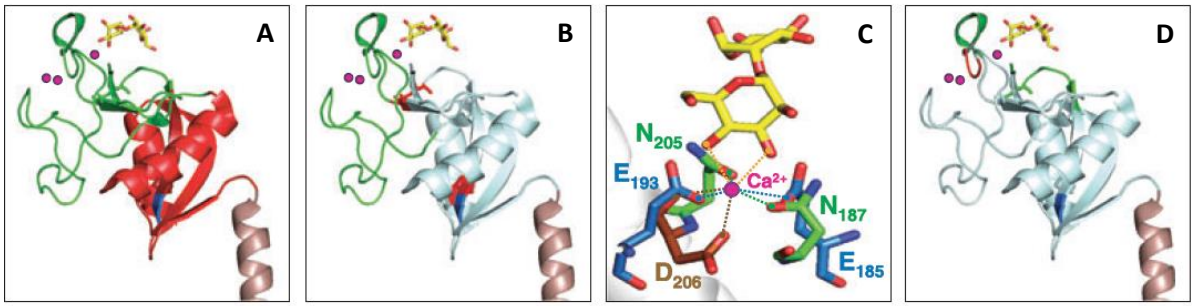


Figure 1-4 Crystal structure of the CRD. Panel A highlights the two lobes of the CRD fold, shown in red and green; Panel B displays the disulphide bridges, shown in red; Panel C demonstrates the carbohydrate binding site and interactions of key residues; and panel D shows the two binding-groove regions of the CRD in green. Image adapted from Veldhuizen *et al*., 2011.

Up to four calcium-binding sites can be identified across the collectin family, which differ in their relative affinity for the metal ion. Location of calcium sites 1, 2 and 3 is considered to be crucial for the stability of the domain, as removal of these sites increases the protein's susceptibility to protease degradation and conformational change (Cooley *et al.,* 2008); while site 4 is directly involved in saccharide binding and the calcium-dependency of collectin ligand binding.

Collectins can be divided into two groups based on their specificity for either mannose or galactose. With the exception of CL-P1, all the described collectins have a binding preference for

mannose-type ligands over galactose-type. SP-A, SP-D and MBL have also been shown to have even broader binding specificity, including nucleic acids, phospholipid and nonglycosyalted proteins (Palaniyar *et al.,* 2004; Kuroki & Sano, 1999; Ip *et al.,* 2009). Studies of the preference of simple monosaccharides provides a general understanding of binding specificity of the collectins, and despite the overall similarities, it becomes clear that additional structural factors are involved; crystallographic studies have revealed some of the underlying mechanisms of binding by the CRD and highlighted involvement of several key residues in the binding pocket.

The majority of studies concerning collectin binding have so far focused on binding to monosaccharides representative of terminal carbohydrate residues. It has been shown that multiple CRDs can bind simultaneously to monosaccharide units of different polysaccharide chains and computational docking studies suggest that amino acids outside the binding pocket may also be necessary for longer ligand binding by collectins (Allen *et al.,* 2001). SP-D has also been shown to bind longer ligands such as maltotriose and *p*-nitrophenyl maltoside whereby in addition to the terminal carbohydrate binding, a phenylalanine residue outside the binding site interacts with the third saccharide (Crouch *et al.,* 2006).

In addition to the surface structure of the collectins, multimerisation of oligomers may also influence multivalent binding to ligands (Lee & Lee, 2000). Although the CRD structures of collectins are very similar, the actual trimeric structures vary significantly. For instance, the angle between the neck and lectin domain of SP-A is close to 90°, creating a flat, T-shaped trimer, whereas a larger kink between the head and neck domains in SP-D and MBL results in a Y-shaped trimer (see Fig 1-5).

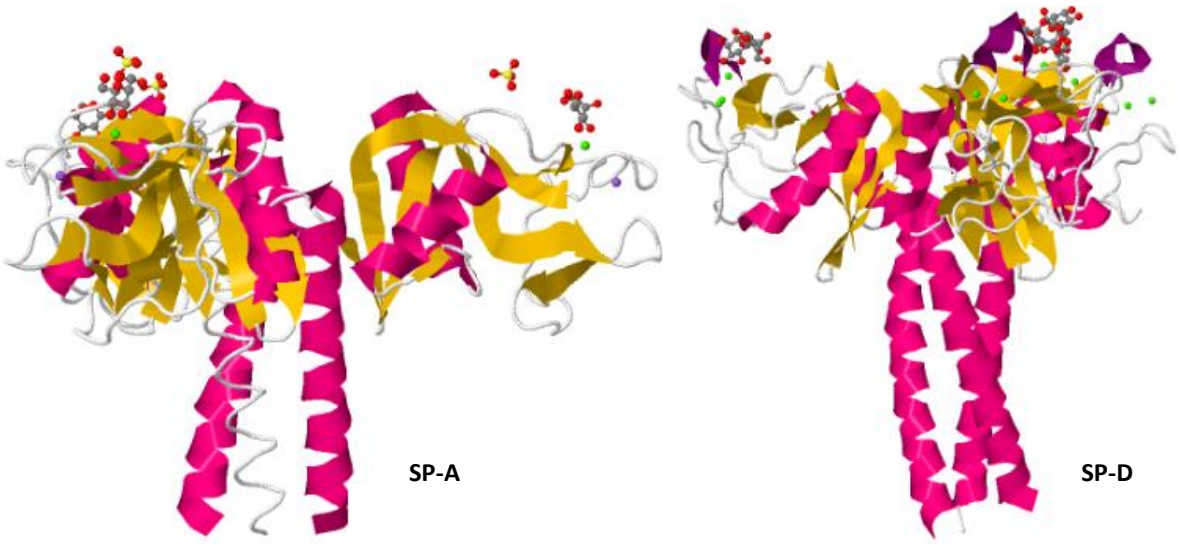


Figure 1-5 Crystal structures of trimeric SP-A and SP-D in side-view. The CRD domain is shown in yellow and white and calcium ions are shown as green spheres. Coordinates taken from the Protein Data Bank: accession numbers 3PAK (left) and 3G83 (right).

Spatial positioning of CRDs in the collectins affects the array of monosaccharides the protein can bind to - and the pattern of oligosaccharides on an invading pathogen must fit the spatial organisation of binding sites for the specific collectin (Seaton *et al.*, 2010).

1.1.4 Receptors for collectins

In addition to recognition and binding of physiological ligands on the surface of invading pathogens, collectins display specific interactions with host cells. A crucial function of the collectins lies in their ability to enhance phagocytosis and induce opsonisation, which in most cases is mediated by calcium-dependant binding to microorganisms. The following table summarises the binding of collectins MBL, SP-A and SP-D to cell surface receptors and the biological consequences that ensue:

Table 1-3 Binding of collectins to cell surface receptors and the biological consequences. Haczku, 2008.

Receptor	Cells	Collectin	Binding site	Consequence
CR1 (CD35, complement receptor 1)	B-cells, monocytes, neutrophils	SP-A, MBL	Collagen tail	Phagocytosis

Table 1-4 Binding of collectins to cell surface receptors and the biological consequences. Haczku, 2008.

Receptor	Cells	Collectin	Binding site	Consequence
Calreticulin-CD91 complex	Macrophages, neutrophils	SP-A, SP-D, MBL, conglutinin, CL-43	Collagen tail	All: apoptotic cell clearance SP-A and SP-D: pro-inflammatory cytokine release, p38 and NF-κβ activation
SP-R210 (SP-A receptor 210)	Macrophages, T-cells	SP-A	CRD	Phagocytosis, increased nitric oxide and TNFα production, reduced T-cell proliferation
CD14	Myeloid lineage cells	SP-A, SP-D, MBL	SP-A: neck region SP-D: CRD MBL: CRD-independent	Reduced cytokine production
TLR2 (Toll-like receptor 2)	Alveolar macrophages	SP-A, SP-D	CRD	Reduced TNFα production
TLR4 (Toll-like receptor 4)	Macrophages	SP-A, SP-D	CRD	Increased NF-κβ activation
SIRPα (signal- inhibitory regulatory protein α)	Macrophages	SP-A, SP-D	CRD	Reduced cytokine production

1.2 Functions of pulmonary Surfactant Protein-D

Pulmonary surfactant is a mixture of lipids (90%) and proteins (10%) that establishes the liquid phase of the alveolar epithelium and maintains surface tension to prevent lung collapse. Surfactant is stored as lamellar bodies - then secreted into the alveolar spaces and airways where it forms a lattice, referred to as tubular myelin. This is considered to be an intermediate product

that forms the lipid film that coats the alveoli. Surfactant protein-D was originally identified as a surfactant-associated protein responsible for pulmonary surfactant homeostasis and constitutes 0.6% of surfactant proteins (Weaver & Whitsett, 1991; Kishore *et al.,* 2006). Subsequent structural and functional characterisation have since proved that in addition to this, SP-D is a potent innate immune molecule involved in a wide range of immune responses, such as recognition and clearance of viruses, bacteria, fungi, apoptotic and damaged cells, downregulation of allergic reactions to pollen and dust-mites, as well as orchestration and modulation of the innate and adaptive immune systems (Kishore *et al.,* 2006; Cormack & Whitsett, 2002).

SP-D is a large hydrophilic protein that is secreted by by alveolar type II and non-ciliated bronchiolar epithelial cells as dodecamers consisting of four collagenous trimers cross-linked by disulphide bonds (Kuan *et al.,* 1992).

1.2.1 Biosynthesis and genetic organisation

Crouch and co-workers characterised cDNA specific for human SP-D and demonstrated that the SP-D gene has a coding sequence spanning >11kilobases, situated on the long-arm of chromosome 10. Sequencing studies have shown that the signal peptide/amino-terminus, neck domain and CRD are each encoded by a single exon, like those of SP-A and MBL. However, a unique split-intron-exon structure is present for the collagen region which is encoded by five exons, including four tandem exons. The latter exons reveal a marked conservation in the structure of the collagen domain, consistent with replication of this sequence during evolution, and distinguishes SP-D from other collagenous C-type lectins. Split sequences may offer certain evolutionary advantages for collagen-containing proteins, since splicing of the genes will reconstitute an intact glycine codon, maintaining the Gly-X-Y triplet (Crouch *et al.,* 1993).

SP-D is extensively modified post-translationally, and mass spectrometric analysis of intact proteins indicate modifications are heterogeneous in nature. Modifications include cleavage of the signal peptide, partial hydroxylation of proline and lysine residues in the collagen

domain to stabilise the triple helix, and assembly in to subunits and high-order oligomers through disulphide linkages and non-covalent interactions. (Leth-Larsen *et al.,* 1999).

Native SP-D is composed of oligomers of a 130kDa subunit comprising three identical polypeptides chains of 43kDa. It is assembled into a 520kDa tetrameric structure with four homotrimeric subunits linked via their N-terminal regions in a cruciform (see Fig 1-1). Electron micrographs of rat SP-D show that the subunits emanate from a central pore into two opposite pairs. Each subunit measured approximately 46nm in length and was terminated by a globular region 8-9nm in diameter. Preparations also showed high-order multimers, referred to as *fuzzy balls*, consisting of SP-D molecules associated at their N-termini (see figure below).

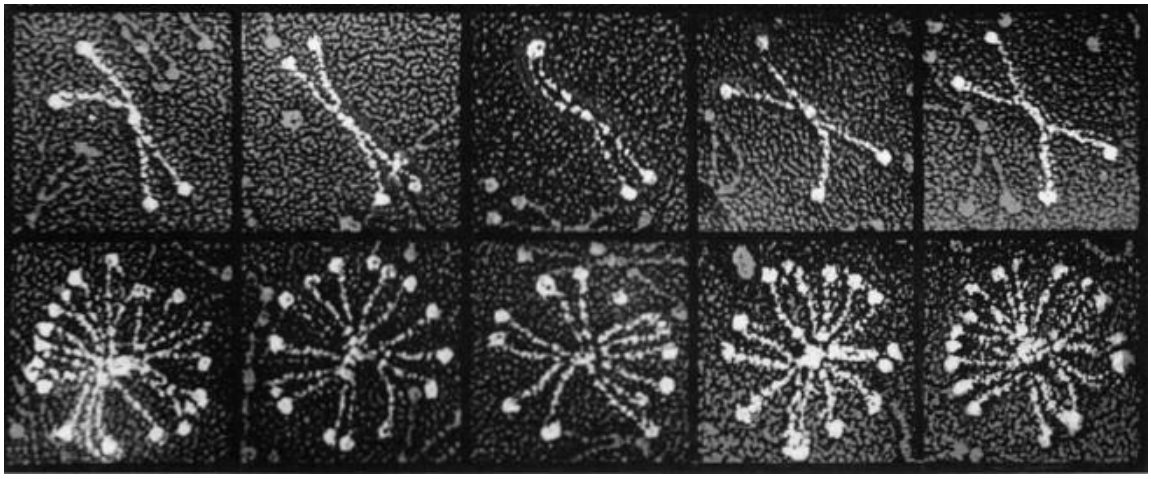


Figure 1-6 Electron micrograph of rat SP-D showing the dominant structure in cruciform assembly (top) and multimers (bottom). Figure adapted from Crouch *et al.,* 1994.

1.2.2 Pathogens recognised by SP-D

One of the first studies indicating surfactant's involvement in pulmonary immunity was completed by LaForce *et al.,* 1973 and demonstrated enhanced killing, but not the phagocytosis of *Staphylococcus aureus* by alveolar macrophages. Although found elsewhere in the body, SP-D's primary location is in bronchoalveolar lavage fluid of the lung where it acts as a first line of defence against inhaled pathogens; the enormous surface area of the lung epithelium is

constantly exposed to particles as large as 5µm in diameter that are deposited in the terminal airways of alveoli (Holgate, 1999). The physiological response to pathogen binding by SP-D has been observed for a wide range of microorganisms and allergens, and is summarised in the table below. In addition to its antimicrobial properties SP-D plays a vital role in the control of inflammatory responses triggered by allergens and clearance of diseases cells. The response to binding of a pathogen includes various innate immune mechanisms, such as agglutination or aggregation to create a clumped mass and inhibit colonisation, enhanced phagocytosis and clearance of the pathogen by opsonisation, increased production of reactive oxygen species by neutrophils and monocytes, recruitment of chemokines and cytokines, and modulation of the immune system (Kishore *et al.,* 2006).

Table 1-5 Interactions of SP-D with pathogens, showing the target of binding and subsequent immune response. Cells labelled 'tbd' are to be determined.

Microbe	Target ligand	Consequence of binding	Reference
Gram-negative bacteria			
Escherichia coli	Lipopolysaccharide	Agglutination, ↑ phagocytosis, growth inhibition	Kuan *et al.,* 1992; Wu *et al.,* 2003
Salmonella minnesota	Lipopolysaccharide	Agglutination, ↑ phagocytosis	Kuan *et al.,* 1992; Bufler *et al.,* 2003
Haemophilus influenzae	Lipopolysaccharide	↑ phagocytosis (moderate)	Restrepo *at al.,* 1999
Klebsiella pneumonia (unencapsulated)	Lipopolysaccharide	Inhibits growth	Restrepo *et al.,* 1999
Legionella pneumophila	Lipopolysaccharide	Inhibits growth	Sawada *et al.,* 2010
Pseudomonas aeruginosa	Lipopolysaccharide	↑ phagocytosis	Lim *et al.,* 1994
Helicobacter pylori	Lipopolysaccharide O-antigen	Agglutination	Khamri *et al.,* 2005

Table 1-6 Interactions of SP-D with pathogens, showing the target of binding and subsequent immune response. Cells labelled 'tbd' are to be determined.

Microbe	Target ligand	Consequence of binding	Reference
Gram-positive bacteria			
Bacillus subtilis	*Bacillus subtilis*	*Bacillus subtilis*	*Bacillus subtilis*
Staphylococcus aureus	*Staphylococcus aureus*	*Staphylococcus aureus*	*Staphylococcus aureus*
Streptococcus pneumoniae	Carbohydrate moieties (tbd)	Agglutination, ↑ phagocytosis (failure to enhance killing by neutrophils)	Jounblat *et al.,* 2004
Mycobacteria			
Mycobacterium avium	Lipoarabinomannan	↑ phagocytosis	Kudo *et al.,* 2004
Mycobacterium tuberculosis	Lipoarabinomannan	Agglutination, ↓ phagocytosis	Ferguson *et al.,* 1999
Mycoplasma pneumoniae	Glycolipids (Ca^{2+} independent)	Neutralisation, ↑ phagocytosis, ↑ TNFα production	Chiba *et al.,* 2002, Chiba, 2003
Viruses			
Influenza A	High-mannose glycoproteins haemagglutinin, neuraminidase	Neutralisation, ↑ phagocytosis	Hartshorn *et al.*, 2000
Human immunodeficiency virus	Glycoprotein-120	Neutralisation	Meschi, 2005
Rotavirus (bovine)	VP7 glycoprotein	Agglutination, neutralisation	Reading *et al., 1*998
Respiratory syncytial virus	G protein F protein	Neutralisation	Hickling *et al.,* 1999
SARS coronavirus	Spike glycoprotein	↑ phagocytosis	Leth-Larsen *et al.,* 2007; Wu *et al.,* 2009
Fungi			
Aspergillus fumigatus	Glycoprotein-45 & -55	Agglutination, ↑ phagocytosis	Madan *et al.,* 1997; Madan *et al*., 2001

Table 1-7 Interactions of SP-D with pathogens, showing the target of binding and subsequent immune response. Cells labelled 'tbd' are to be determined.

Microbe	Target ligand	Consequence of binding	Reference
Blastomyces dermatitidis	β-glucan	↑ phagocytosis, ↓TNFα production	Lekkala *et al.,* 2006
Candida albicans	Mannan polysaccharides	Agglutination, growth inhibition, ↑ phagocytosis	Awasthi *et al.,* 2004; Van Rozendaal *et al.,* 2000
Coccidioides posadasii	Mannose-containing Coccidiodal antigens (tbd)	Surfactant modification, ↑ phagocytosis	Ampel *et al.,* 2005; Awasthi *et al.,* 2004;
Cryptococcus neoformans	Glucuronoxylomannan, mannoprotein-1	Agglutination, ↑ phagocytosis	Awasthi *et al.,* 2004; de Wetering *et al.,* 2004b
Histoplasma capsulatum	Glucuronoxylomannan, mannoprotein-1	Agglutination	Schelenz *et al.,* 1995
Pneumocystis jiroveci	Surface glycoprotein-A, β-glucan, glycoprotein-120	Agglutination, attachment to macrophages	Yong *et al.,* 2003
Saccharomyces cerevisiae	β-glucan	Agglutination	Allen *et al.,* 2001
Protozoa			
Schistosoma mansoni	Fucosylated glycoconjugates	↑ phagocytosis	de Wetering *et al.,* 2004b
Pollen	Carbohydrate structures on pollen starch granules (tbd)	↑ phagocytosis, ↓ allergic hypersensitivity	Currie *et al.,* 2000
Dust mites	Glycoprotein (tbd)	↑ phagocytosis, ↓ allergic hypersensitivity	Erpenbeck, 2005; Strong *et al.,* 2003

1.2.2.1 Gram-negative bacteria

Gram-negative bacteria are a class of prokaryotic microorganisms that do not retain crystal violet when Gram-stained, due to a thin peptidoglycan layer in the cell wall. They can cause infections including pneumonia, bloodstream infections, wound or surgical site infections and meningitis, and are becoming increasingly resistant to antibiotics (cdc.gov). Numerous in vitro studies have shown that SP-D increases receptor-mediated uptake of a variety of bacteria by enhancing phagocytosis via opsonisation, but also directly by upregulating cell-surface receptors on macrophages independent of microbial binding.

The outer membrane of Gram-negative bacteria consists predominately of lipids which make up ~75% of the total membrane surface, attached to which are long chain polysaccharides.

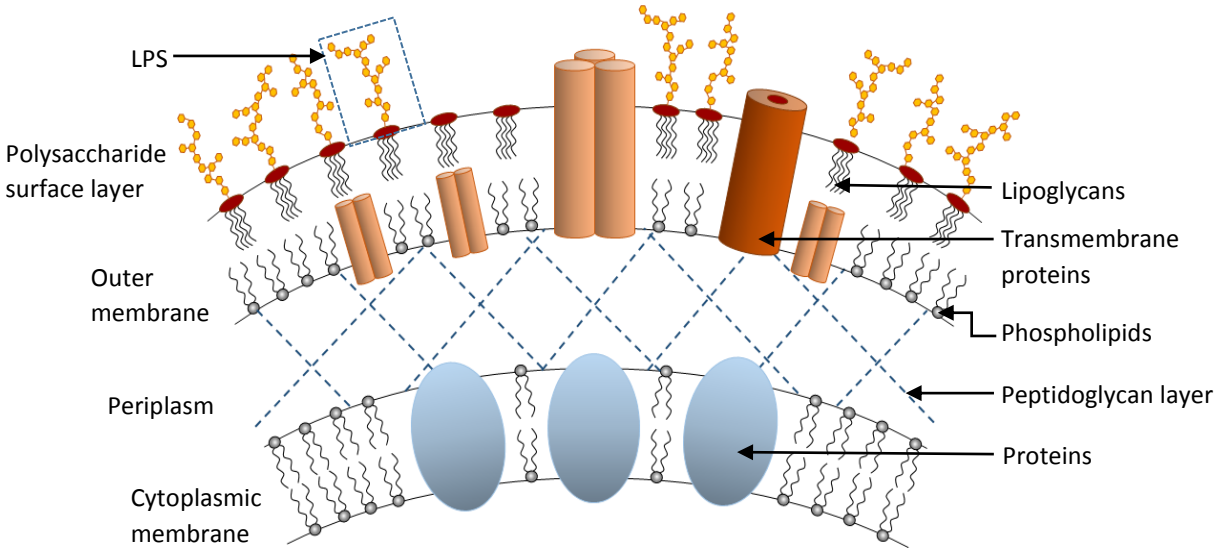


Figure 1-7 The cell wall of Gram-negative bacteria, showing the cytoplasmic membrane (bottom, internal to the bacterium), the periplasm consisting of a thin peptidoglycan layer, and an outer membrane to which lipopolysaccharides are anchored - a single LPS molecule is shown in the hashed box. Alexander & Rietschel, 2001.

1.2.2.1.1 Lipopolysaccharide inner core

Lipopolysaccharide (LPS), also known as endotoxin, is an established SP-D ligand and constitutes a major component of the cell wall of Gram-negative bacteria. In 1884 Richard Pfeiffer and Robert Koch discovered that lysates of cholera-inducing bacteria caused toxic shock in guinea pigs, and postulated that these heat-stable toxins were located inside the bacterial cell, hence the pseudonym *endotoxin* (Brade, 1999). It exists on the outer layer of the membrane and is, in noncapsulated strains which contain no polysaccharide cell-envelope, exposed on the cell surface - among them several human pathogenic species such as *Escherichia coli*, *Salmonella enterica*, *Neisseria meningitidis*, *Haemophilus influenzae*, *Bordetella pertussis*, *Pseudomonas aeruginosa*, *Helicobacter pylori*, *Klebsiella pneumoniae*, *Legionella pneumophila* and *Chlamydia trachomatis* (Alexander & Rietschel, 2001).

Much of the understanding of LPS biosynthesis and structure is founded on characterisation of *E. coli* and *Salmonella*, which demonstrated that LPS is composed of a hydrophobic region (lipid A) covalently linked to polysaccharide. The latter can be divide into two groups: an inner core region proximal to the lipid A; and an outer core that provides a site of attachment for longer O-polysaccharides.

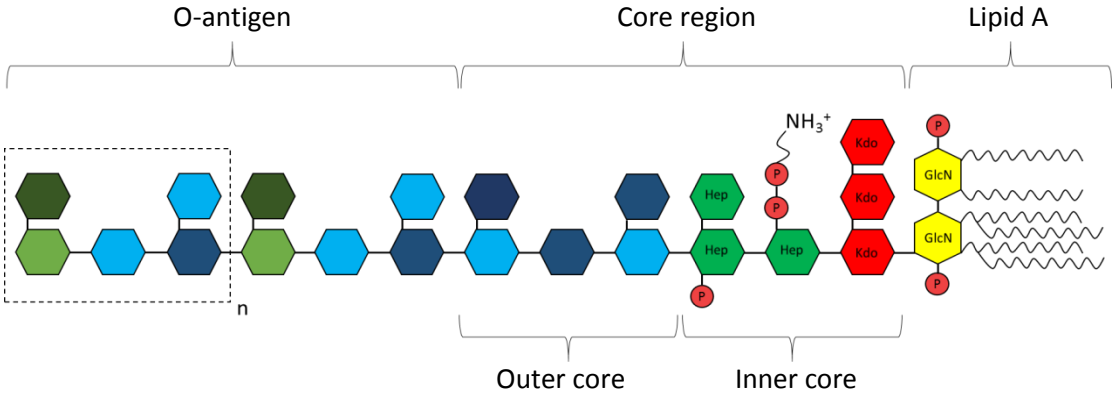


Figure 1-8 The general structure of LPS showing hydrophobic region (lipid A), core region made up of an inner and outer portion, and the O-antigen. The terminal structure (hashed box) is the repeating unit of the O-antigen. GlcN - glucosamine; Kdo - 3-deoxy-D-manno-octulosonic acid; Hep - D-glycero-D-manno-heptose. Image adapted from Alexander & Rietschel, 2001.

The toxicity of LPS is associated with the lipid component, lipid A, which is understood to be a potent activator of innate immunity (Raetz & Whitfield, 2002). It exists as a glucosamine-based phospholipid situated in the outer membrane of the bacterial cell surface of most Gram-negative bacteria. A single *E. coli* bacterium contains roughly 10^6 lipid A residues, and a minimal number are required for growth. Given their conserved architecture, most lipid A molecules are detected at picomolar levels by the innate immune system by receptors such as TLR-4 present on macrophages. Activation triggers biosynthesis of a range of inflammatory mediators, such as upregulation of TNFα and interleukin-1 or stimulation of the adaptive immune response (Dinarillo, 1991; Werling & Jungi, 2003). When overproduced, lipid A can cause severe sepsis - Gram-negative septic shock can be accompanied by disseminated intravascular coagulation and multiple organ failure (Engelmann & Massberg, 2014). In a case that gained public interest in 2011, an *E. coli* outbreak in Germany caused haemolytic uraemic syndrome and the death of 31 people. The first and largest bioterrorism attack in United States history involved infection of salad bars with *Salmonella typhimurium* in which 751 people contacted salmonellosis as a result (Tucker, 1999).

The structure of the inner core is well conserved within a bacterial genus, and as distantly related bacteria share structural features, this reflects its importance in membrane integrity. The inner core typically consists of Kdo and heptose residues and is often decorated with additions of other monosaccharides or phosphate-containing residues such as phosphoethanolamine and phosphorylcholine.

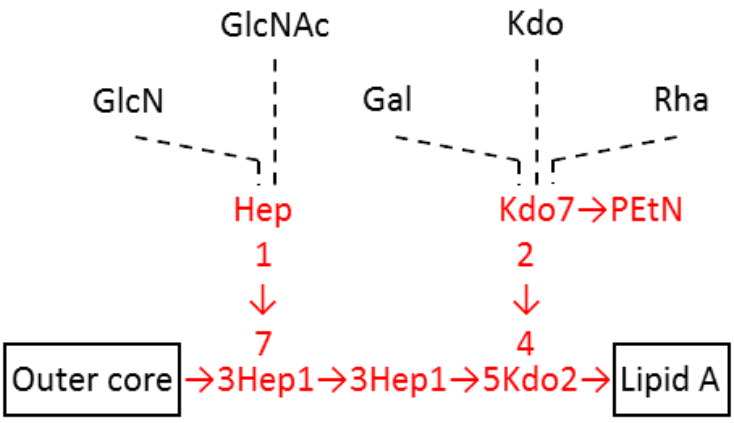


Figure 1-9 The structure of lipopolysaccharide inner core region from *E. coli* and *Salmonella*. Residues and bonds in red are conserved, while type-specific additions are shown by hashed lines. Hep - L-glycero-D-manno-heptose, Kdo - 3-deoxy-D-manno-octulosonic, GlcN - glucosamine, GlcNAc - N-acetylglucosamine, PEtN - phosphoethanolamine, Gal - galactose and Rha - rhamnose. All bonds in α-conformation. Adapted from Heinrichs *et al.,* 1998.

The inner core structure described in Figure 1-8 is conserved across a wide range of bacterial genera including *Bordetella*, *Burkholderia*, *E. coli*, *Klebsiella*, *Providencia*, *Salmonella*, *Serratia*, *Shigella* and *Yersinia* species (Caroff & Karibian, 2000; Ortega *et al.,* 2009; Kondakova *et al.*, 2006; Holst, 2011; Helander *et al.,* 1996; Vinogradov *et al.,* 2003; Kondakova *et al.,* 2010; Vinogradov *et al.,* 2002). Similarities exist between both *Neisseria* and *Pseudomonas* cores which contain only the first two, inner heptose residues (Cox *et al.,* 2003; Bystrova *et al.,* 2004) and *Haemophilus*, *Helicobacter*, *Pasteurella* and *Vibrio* species, which differ only in that they contain an α1-2-linked outer heptose residue (Orgeig *et al.,* 2010; Holst, 2011). Members of the *Chlamydia* family contain a unique LPS structure, expressing a Kdo triplet only (Erridge *et al.,* 2002).

Modification of the core with phosphates contributes to membrane integrity and is required for transport of LPS and transport to the outer membrane. Expression of a mutant strain of *Psuedomonas aeruginosa* that lacked the enzyme responsible for phosphorylation showed halted growth, an accumulation of saccharides inside the cell and disruption of the outer membrane permeability barrier (DeLucia *et al.,* 2011).

The understanding of *E. coli*, *Salmonella* and *Klebsiella* biosynthetic genes involved in core assembly is aided by the clustering on a single chromosome. This locus, termed *waa*, encodes all the transferases required for inner core assembly and genomes of distantly related bacteria show similar clusters of genes. *Neisseria* and *Haemophilus* show no significant clustering of *waa*-type genes, which may explain the difference in structure of these two species (see Fig 1-10). Waa enzymes involved in the assembly of the inner core include WaaA, a bifunctional Kdo transferase responsible for addition of one or more Kdo residues to the lipid A moiety; WaaC, a heptosyltransferase responsible for the addition of the first heptose to Kdo; WaaF that adds the second heptose to the first; WaaP, a kinase that phosphorylates the first heptose; WaaQ that adds a third heptose to the second; and WaaY that phosphorylates the second heptose.

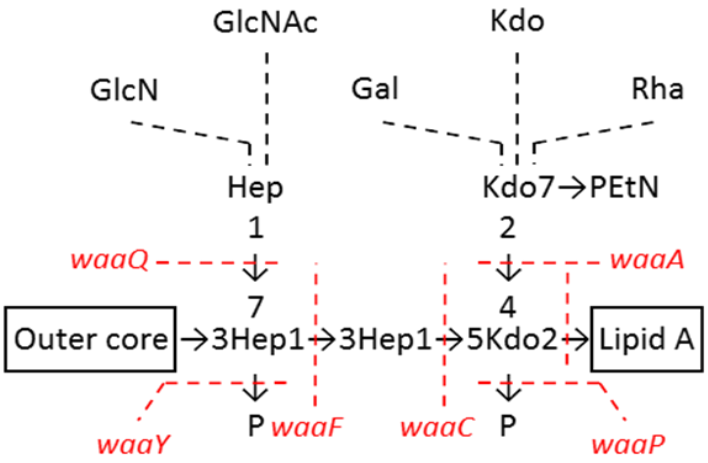


Figure 1-10 Structure and biosynthesis of the inner core from *E. coli* and *Salmonella*. *Waa* transferases that form the core are shown in red, the dashed lines indicate the position in the structure to which specific gene products are added. All bonds in α-conformation. Adapted from Heinrichs *et al.*, 1998.

1.2.2.1.2 Lipopolysaccharide Outer core

The outer core of LPS shows a greater structural diversity than the inner, and reflects the increased exposure of this region to the pressures of the immune response. For example, there

are five known outer core types in *E. coli*: R1, R2, R3, R4 and K-12, and each utilises similar Waa enzymes depending on the type-specific residues in the structure.

The additional Waa transferases required for assembly of the outer core vary according to bacterial subtype. For instance, *E. coli* R1 cores employ WaaG which adds a glucose to the outermost heptose of the inner core; WaaO, adds a second glucose; WaaV, adds a third glucose; WaaT, which adds a galactose residue to the second glucose; WaaW, adds a second galactose to the first; and WaaL, which ligates the polysaccharide to an O-antigen (See Fig 1-11).

It has been suggested that the outer core of Gram-negative bacterial LPS is required for bacterial entry into epithelial cells of the host. It was demonstrated that the outer-most glucose residue in the chain was crucial for recognition and internalisation by the polar membrane of epithelial cells (Hoare *et al.,* 2006).

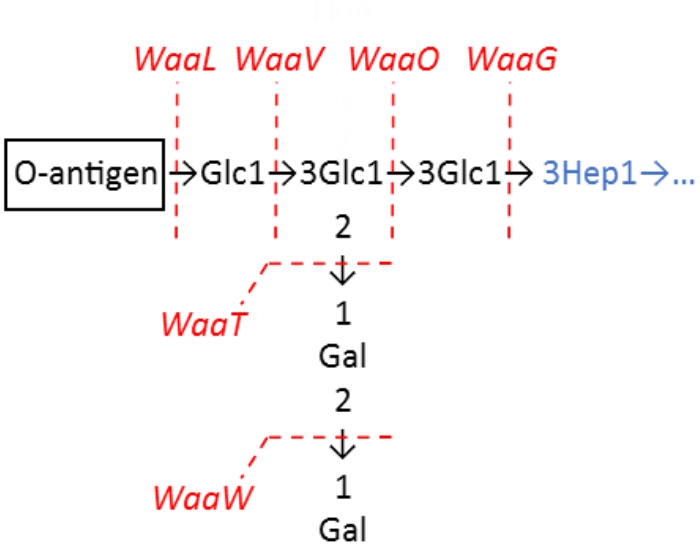


Figure 1-11 Structure and biosynthesis of the outer core from *E. coli* R1. The adjoining outermost heptose that links the inner and outer cores is shown in blue. *Waa* transferases that form the core are shown in red, the dashed lines indicate the position in the structure to which specific gene products are added. All bonds in α-conformation. Adapted from Heinrichs *et al.,* 1998.

The structure of the *Salmonella* outer core shows similarities to the *E. coli* types, and the same *waa* gene types are responsible for its assembly.

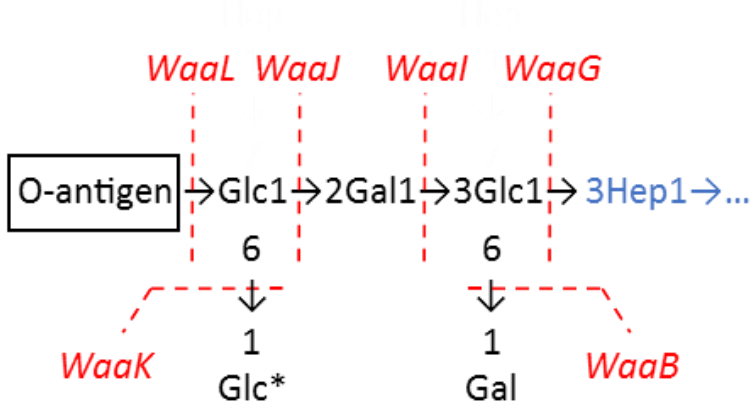


Figure 1-12 Structure and biosynthesis of the outer core from *Salmonella*. The adjoining outermost heptose that links the inner and outer cores is shown in blue. *Waa* transferases that form the core are shown in red, the dashed lines indicate the position in the structure to which specific gene products are added. The residue denoted by * can be glucose or N-acetylglucosamine. All bonds in α-conformation. Adapted from Heinrichs *et al.*, 1998.

Mutant strains deficient in one or more of the aforementioned Waa enzymes are termed *rough mutants*, and show truncated inner and outer cores.

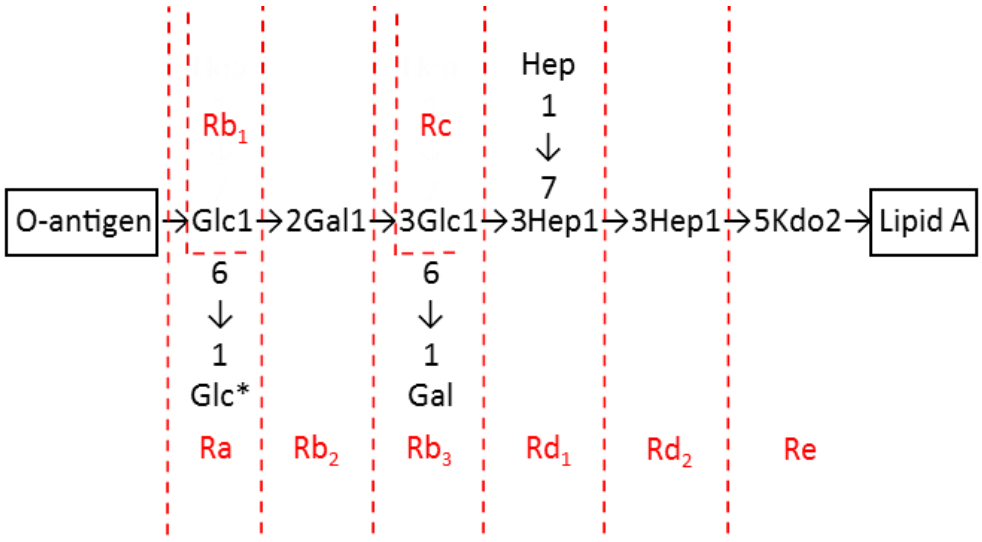


Figure 1-13 The structure of rough mutants in the core region of *Salmonella* LPS. Truncation ranges from the shortest mutant, Re, that contain Kdo residues only, to Ra mutants that contain the whole inner and outer core structures. The residue denoted by * can be glucose or N-acetylglucosamine. Adapted from Mansfield *et al.*, 1996.

1.2.2.1.3 Lipopolysaccharide O, K and H-antigens

Lipopolysaccharide can be described as either rough or smooth depending on the presence or absence of an O-antigen. Additional core and O-antigen polysaccharides, although not vital for growth, enable bacteria to resist antibiotics and evade the immune system. Evasion of the immune system by molecular mimicry has been observed in strains of *Helicobacter pylori*, whereby the O-antigen expresses $Lewis_x$ and $Lewis_y$ antigens identical to those occurring in the human gastric mucosa (Appelmelk *et al.,* 1997). In the vast majority of LPS structures, the O-antigen chain is characterised by an extremely high structural variability and is comprised of a repeating oligosaccharide unit of two to six saccharides. The distinctive O-antigen structures are used to allocate serogroups to bacteria and over 160 serogroups in *E. coli* alone have been assigned (Stenutz *et al.,* 2006).

In addition to O-groups, other antigens exist on the pathogen cell surface: capsular polysaccharides of Gram-negative bacteria, also called K-antigens, maintain structural integrity of the cell and contribute significantly to virulence. They form a defensive barrier against host immunity and grant the ability of bacteria to adhere to epithelial cells as well as other bacteria (Aduse-Opoku *et al.,* 2005). Furthermore, flagella of motile Gram-negative bacteria also contribute to bacterial virulence, chemotaxis and invasion of host cell. Composed of the protein flagellin, they are post-translationally modified through addition of various glycosylation sites. Families of Gram-negative bacteria known to contain flagella include *E. coli*, *Helicobacter*, *Pseudomonas*, *Salmonella* and *Vibrio*, and the antigenic moieties present on flagella are termed H-antigens (Ni *et a*l., 2008). Both capsular polysaccharide and flagellin have been shown to induce upregulation of, and be bound by SP-D.

1.2.2.2 Gram-positive bacteria

Gram-positive bacterial infections are a leading cause of pneumonia, a significant source of morbidity and mortality throughout the world. Of the pathogens that cause pneumonia,

Streptococcus pneumoniae and *Staphylococcus aureus* are the most common, both of which have been shown to be recognised by SP-D (Osiyemi & Dickinson, 2000; de Wetering *et al.,* 2001).

Unlike Gram-negative, Gram-positive bacteria retain crystal violet stain due to a thick peptidoglycan layer on the cell wall. They lack an outer membrane and as such tend to be more receptive to antibiotics, and in addition they express no lipopolysaccharides. Instead, the two major cell wall components expressed by Gram-positive bacteria are peptidoglycan (PepG) and Lipoteichoic acid (LTA) which can induce inflammation (Percy & Gründling, 2014). SP-D has been shown to bind both LTA and PepG components of Gram-positive bacteria (de Wetering *et al.*, 2001).

1.2.2.2.1 Peptidoglycan

Peptidoglycan, PepG, consists of alternating β-linked N-acetylmuramyl and N-acetylglucosaminoglycan residues, cross-linked by short peptides. Bacteria are encapsulated by multiple layers of cross-linked PepG, and it is therefore the most prevalent component of the cell wall - accounting for up to 40% of the weight of the bacterial cell (Shockman & Barrett, 1983).

1.2.2.2.2 Lipoteichoic acid

Lipoteichoic acid, LTA, is a member of the amphiphile family of glycolipids composed of a hydrophobic diacylglycerol membrane anchor and a hydrophilic side chain that varies across species. They usually consist of glycerophosphate repeats substituted with D-alanine or glycosyl groups.

1.2.2.3 Mycobacteria

Mycobacterium tuberculosis is a highly pathogenic bacterial species of the family *Mycobacteriaceae*, and the causative agent of tuberculosis; *M. avium* is an opportunistic bacterium that causes localised and disseminated infection of the lung, bone and intestine, common in immunocompromised individuals. They contain a unique, waxy coating on its cell

surface that renders it impervious to Gram staining (Trifiro *et al.,* 1990). This layer is composed of peptidoglycan, peripheral lipids and surface glycolipids called *lipoarabinomannan* – mannosylated moieties that act to suppress the immune response (Carlson *et al.,* 2009).

1.2.2.4 Viruses

Many viruses have been shown to be recognised and neutralised by SP-D. Influenza viruses are RNA viruses of the family *Orthomyxoviridae*, and are divided in to genera A, B and C. Influenza A viruses (IAV) are further classified by the presence of proteins on their cell surfaces: Haemagglutinin (HA) and Neuraminidase (NA) and are known to infect humans and other mammals, vertebrates and birds. The only subtypes currently circulating in humans (designated by their HA and NA proteins) are H1N1, H3N2 and H5N1.

The HA glycoprotein is expressed on the viral membrane as a single polypeptide that is cleaved by host trypsin-like proteases to produce two peptides HA_1 and HA_2 in order to become infectious. Once cleaved, a newly exposed N-terminus of the HA_2 peptide is able to fuse the viral envelope to cellular membranes of host cells via sialic acid residues. The sialic acid binding site is a crucial component of IAV, and strains vary in their affinity - determining the animal species which they are able to infect. For instance, avian influenza strains such as H5N1 preferentially bind to sialic acids bound to galactose via an α2,3 linkage, a major sialic acid present on epithelial cells of avian digestive systems. Human IAVs show preference for α2,6-linked residues, consistent with those found on the epithelial surface of the human respiratory tract. Interestingly, cells of the pig trachea express both α2,3 and α2,6-linked sialic acids, and as such can be infected with both human and avian influenza viruses. They are thus considered the 'mixing vessel' for the emergence of new viruses (Olufsson *et al.,* 2005). Carbohydrate addition by IAV has both positive and detrimental effects on virulence; activation by protease cleavage is required for entry to the host cell, but the HA molecule is exposed to the immune system - if carbohydrate moieties on the

protein are in close proximity to the sialic acid binding site they may become blocked, rendering the virus neutralised (Vigerust *et al.,* 2007).

Rotaviruses are encapsulated, double-stranded RNA viruses of the family *Reoviridae* and cause acute viral gastroenteritis. They are unusual in that an expressed glycoprotein, VP7, acts as one of two capsid proteins; VP7 forms a smooth surface on the virus, whereas the other, VP4, protrudes as spikes (Reading *et al*., 1998). VP4 acts as a host-cell attachment protein and viral haemagglutinin (Kalica *et al.,* 1983). The glycans of VP7 are all of the high-mannose type.

Respiratory Syncytial Virus is a negative-sense, single stranded RNA virus of the family *Paramyxoviridae* that causes respiratory tract infections. They express G proteins which contain O- and N-linked oligosaccharide chains, typically high-mannose N-linked polysaccharides (Wertz & Lichtenstein, 1988). Protein N-glycosylation was analysed and illustrated that the carbohydrate moieties are complex-oligosaccharides (see Fig 1-14), comprised of fucose, glucosamine and galactosamine, with a mannose core (Rixon *et al,* 2002). S proteins are membrane glycoproteins that contain a single N-linked glycosylation site that although not fully elucidated, is believed to be a high-mannose oligosaccharide (Zhou *et al*., 2010).

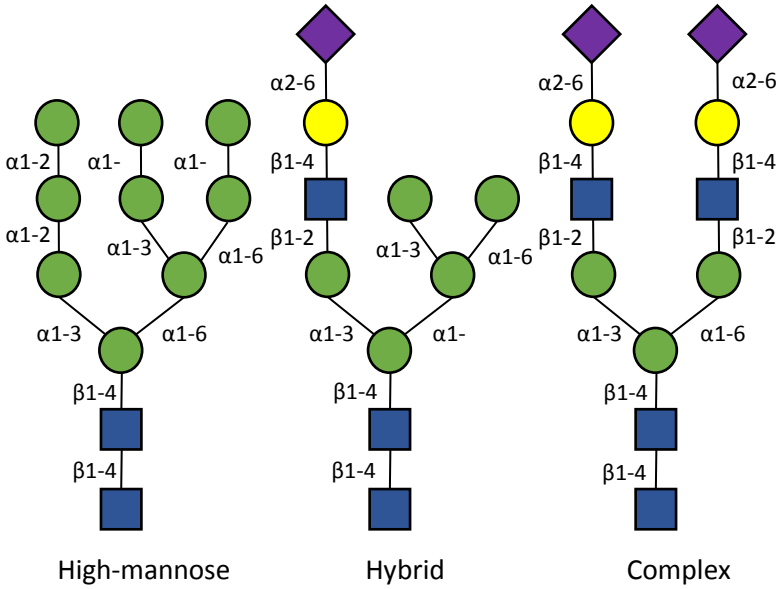


Figure 1-14 High-mannose (left), hybrid (centre) and complex oligosaccharides (right). Mannose (green), GlcNAc (blue), galactose (yellow) or fucose residues (purple). Adapted from Sigma-Aldrich, 2015.

1.2.2.5 Fungal pathogens

Fungi are a large family of eukaryotic organisms that include yeast, moulds and mushrooms, some of which are known to be neutralised by SP-D. Opportunistic fungal pathogens represent a significant health burden, especially in immunocompromised individuals. *Cryptococcus neoformans* and *Histoplasma capsulatum* express capsular polysaccharide glucuronoxylomannan (consisting of mannose, xylose and glucuronic acid), and mannoprotein-1 - a protein of about 100 kDa attached to β1-6-glucan through a glycosylphosphatidylinositol anchor, containing five alpha-linked mannosyl residues (Richardson, 1993; Kollar *et al.,* 1997). *Blastomyces dermatitidis* and *Pneumocystis jirovecii* specifically express β1-6-glucan.

1.2.2.6 Schistosomes

Schistosoma mansoni, known as blood flukes, are parasitic flatworms responsible for schistosomiasis infections - considered by the World Health Organisation as the second most socioeconomically devastating parasitic disease after malaria (Who.int, 2015). Adult *S. mansoni*

reside in the veins of mammalian hosts, where they produce eggs that penetrate the digestive tract. When these eggs are excreted, they develop into miracidia which infect freshwater snails. The snail then releases motile cercaria larvae that are able to infect humans by penetration of the skin, where they migrate to the heart and lungs (Crabtree & Wilson, 1986).

S. mansoni express fucose-containing glycoconjugates on their cell surface, which like those of *Helicobacter* strains of Gram-negative bacteria, contain Lewis$_x$ antigens (blood group antigens displayed on the terminus of glycolipids present on the cell surface, see Fig 1-15) identical to those occurring in the human gastric mucosa (Appelmelk *et al.,* 1997). Details of the SP-D binding mechanism are detailed in subchapter 1.4.5.

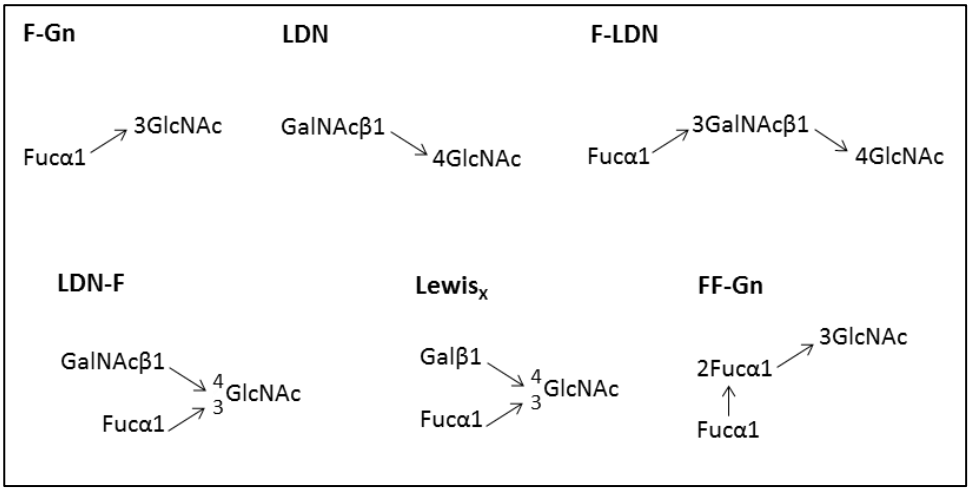


Figure 1-15 Schematic representation of the fucose-containing glycoconjugates of *Schistosoma mansoni* tested for SP-D binding. F-Gn - Fuc(α1–3)GlcNAc; LDN - GalNAc(β1–4)GlcNAc; F-LDN - Fuc(α1–3)GalNAc(α1–4)GlcNAc; LDN-F - GalNAc(β1–4(Fucα1–3))GlcNAc; LewisX - Gal(β1–4(Fucα1–3))GlcNAc; FF-Gn - Fuc(α1–2)Fuc(α1–3)GlcNAc. de Wetering *et al.,* 2004b.

1.2.3 Surfactant homeostasis

Along with preventing pathogen entry into the body, pulmonary surfactant is essential for gas exchange in the lung, and deficiencies in both quality and quantity are associated with respiratory distress. Levels are regulated during synthesis, secretion, reuptake by epithelial cells and degradation by alveolar macrophages to maintain a steady concentration. It has been suggested

that the effect of SP-D on uptake of pathogens may influence macrophage function and signalling, and the role of SP-D in surfactant homeostasis was identified using SP-D-null (-/-) genetically altered mice and by comparing their host defence to mice with the SP-D gene present (±) (Botas *et al*., 1998).

The absence of SP-D caused no structural abnormalities in the lung as a whole and no alterations to the airway epithelial cells was observed. However, abnormalities were seen in the lung tissue itself - enlarged alveoli and increased bronchus-associated lymphocytic tissue (lymphoid tissue that is formed upon inflammation and infection (Randall, 2010)) were consistently detected. Also present were numbers of large, foamy alveolar cells, which appeared to be alveolar macrophages with abundant cytoplasmic vesicles. In tested alveolar lavage fluid, macrophage concentration was 4-fold higher and the mean diameter of cells was 18.75µm compared to 11.75µm in normal, SP-D± mice. Phospatidylcholine levels were 3-fold higher in -/- mice, and large masses of lipids were found to aggregate in the lung.

The results suggest a clear role for SP-D in maintaining phospholipid structure and preventing accumulation of lipid in the lung. It remains unclear whether changes to macrophage numbers and size reflect changes to the surfactant itself - increased accumulation of surfactant may cause increased uptake by macrophages - or represent a primary abnormality as a result of SP-D absence. Lipid aggregation most likely represents the crucial role of SP-D in lipid regulation, but it is unclear how phospholipid homeostasis is modulated, specifically at the level of synthesis and/or catabolism by alveolar epithelial cells or by macrophages (Korfhagen *et al.,* 1998).

1.2.4 Lung hypersensitivity and inflammation

The role of SP-D in development of allergen-induced airway inflammation involves an immune response that reduces inflammation. Levels progressively increase up to 48 hours after allergen challenge, reducing progression of eosinophilia and hypersensitivity. SP-D's interaction with house dust mite and pollen starch granules were demonstrated by administering of intranasal SP-D in a

murine model of pulmonary hypersensitivity. Binding to carbohydrate moieties and glycoproteins of the allergens in a calcium-dependent manner was observed, and resulted in the blocking of IgE-mediated histamine release by stimulated basophils. Overall levels of IgG and IgE were lowered in serum, peripheral and pulmonary eosinophilia was reduced, and allergic Th2 T-helper cells were polarised to protective Th1 phenotypes. Th2 cells subsequently lowered levels of proinflammatory IL-4 and IL-5, while elevating IFN-γ - which influences macrophage function during infection (Madan *et al.,* 2001; Schroder *et al.,* 2003).

1.2.5 Extra-pulmonary presence

Several studies have provided evidence of SP-D expression in tissues other than the lung. 18 extra-pulmonary tissues were tested by Madsen *et al.,* 2000 for the presence of SP-D, and significant levels were detected in the human trachea, brain, testes, salivary glands, heart, prostate, kidneys, small intestine, pancreas and placenta. Weaker expression was seen in the uterus, stomach, mammary glands, spleen, adrenal glands and liver. SP-D expression in epithelial cells of exocrine glands and loops of Henle, ureter and bladder of the urinary tract suggests a major role in the muscosal defence system against invading pathogens. Because SP-D colocalises with gp340 (a glycoprotein receptor reported to be responsible for SP-D membrane binding) in collecting ducts of the kidney, as well as the lung, salivary glands, pancreas and small intestine - it has been suggested that their possible interactions involve processes other than immunprotection (Madsen *et al.,* 2000).

The presence of SP-D in extra-pulmonary environments, especially in organs and tissues most exposed to pathogens, supports its role as an important immune molecule even outside the lung. For instance, the presence of SP-D in the stomach is necessary as this organ is potentially in direct contact with ingested microorganisms. The hydrophobic gut lining bears a strong resemblance to pulmonary surfactant in terms of phosphatidylinositol being the major

phospholipid. More recent studies have found SP-D in other tissues susceptible to infection, such as the paranasal sinus mucosa and Eustachian tube epithelium (Woodsworth *et al.,* 2006; Paananen *et al.,* 2001). The potential role of SP-D in mucosal immunity is yet to be fully understood, but its main function is to prevent microbial colonisation and inhibit growth of pathogens once attached to the epithelium.

It has been proposed that despite its presence elsewhere, the site of primary synthesis of SP-D is the lung. Quick upregulation of serum SP-D in response to acute airway inflammation supports the notion that SP-D is translocated from the airways into the vascular system, in favour of being synthesised systemically (Gaunsbaek *et al.,* 2013).

1.2.6 SP-D in health and disease

Finely-tuned activity of SP-D is vital for a first line of defence against invading pathogens in the lung and elsewhere, and varying levels of both serum and bronchoalveolar lavage are associated with a number of diseases. Physiological levels in the lung undergo changes when challenged by external stimuli such as allergens and asthma, but altered expression in extra-pulmonary locations are also observed. The following table describes a number of disease states and subsequent SP-D levels.

Table 1-8 Changes in SP-D levels in various lung and extra-pulmonary diseases

Disease/disorder	Comments	Reference
Sarcoidosis	SP-D levels increase but vary according to organ.	Kucejko *et al*., 2009; Kitaichi *et al*., 2010
Idiopathic pulmonary fibrosis	Damaged lung epithelium and scar tissue may increase alveolar-membrane permeability	Kuroki *et al*., 1998; Green *et al*., 2001; Barlo *et al.,* 2009
Pulmonary alveolar proteinosis	SP-D accumulates in alveoli and airways, high-order oligomers common	Kuroki *et a*l., 1998; Lin *et al.,* 2008
Tuberculosis	Levels increase but depend on severity of infection	Kondo *et al*., 1998; Kishore *et al*., 2006
Acute lung injury/respiratory distress syndrome	Levels significantly increase at onset but return to normal levels	Greene *et al.,* 1999

Table 1-9 Changes in SP-D levels in various lung and extra-pulmonary diseases

SARS	Levels increase but upregulation of SP-D is inversely proportional to IgG levels	Q. Ashton Acton, 2012; Wu *et al.*, 2009
Smoking	Decreased levels cause increased risk of progression of bronchial dysplasia and lung cancer	Moré *et al.,* 2010; Winkler *et al.,* 2011
COPD	Levels increase during acute attacks	Winkler *et al.,* 2011
Asthma/allergen-induced airway inflammation	Higher levels cause downregulation of IgE production, and supress hypersensitivity	Koopmans *et al.,* 2004; Gemou-Engesaeth *et al*., 2014
Cystic fibrosis	High-order oligomers common	Krane & Griese, 2003; Griese *et al.,* 2004

In addition to levels of SP-D, investigations into the role of SP-D in health and disease have been conducted through the effects of various polymorphisms in gene knock-out mice. Besides reduced pathogen clearance, mice deficient in SP-D show abnormalities in surfactant homeostasis and alveolar cell morphology, and exhibit progressive accumulation of lipids and apoproteins in the alveolar space, hyperplasia of alveolar type II cells, massive enlargement of lamella bodies that store SP-D, and aggregation of foamy alveolar macrophages. These mice spontaneously develop emphysema and pulmonary fibrosis, which suggests a continuous background function of SP-D in otherwise healthy lungs. Lack of SP-D also results in significant lung inflammation when challenged with pathogens compared to wild-type mice, demonstrating SP-D's anti-inflammatory role. Mice also showed substantial increase in broncho-alveolar lavage (BAL) phospholipid and protein levels, as well as increased macrophage content.

1.2.6.1 Pulmonary disorders

Synthesis and secretion of SP-D is increased with injury to the lung. Surfactant abnormalities arise and BAL SP-D levels are raised in lung diseases such as asthma, bronchiolitis, chronic obstructive pulmonary disorder (COPD), lung transplantation, respiratory distress syndrome, pulmonary oedema, sarcoidosis, pulmonary fibrosis, alveolar proteinosis, cystic

fibrosis, pneumonia and infections caused by influenza A viruses or *Pneumocystis* pneumonia (Kuroki *et al.,* 1998). Interestingly, cigarette smoking - which is associated with increased risk of COPD, pneumonia and other lung infections - appears to reduce SP-D levels in otherwise healthy individuals; it is unclear as to whether lowered levels are a cause or consequence of smoking-related diseases (Honda, 1996).

Altered expression of SP-D acts as a useful marker for lung disease, and is considered diagnostic for a number of respiratory infections. For instance, SP-D production is used as a measure of lung maturity in neonates and infants with respiratory distress syndrome, and levels are measured from amniotic fluid and tracheal aspirates respectively (Chida *et al.,* 1997).

A study investigating the role of genetic variants in lung-associated diseases shows that some individuals have increased susceptibility to certain respiratory infections. Genetic polymorphisms of the SP-D gene result in genotypes that show higher association with disorders such as COPD in Mexican populations, tuberculosis in people of Mexican and Indian origin, and RSV infection in people of Finnish descent (Guo *et al*., 2001; Floros *et al.,* 2000; Lahti *et al.,* 2002).

1.2.6.2 Extra-pulmonary disease

SP-D in sites other than the lung also exhibits varying expression patterns in response to disease. For example, attacks of chronic rhinosinusitis, which involved inflammation of the nasal cavities triggered by infection or allergy, result in upregulation of SP-D (Ooi *et al.,* 2007).

Consistent with the surfactant-related function of SP-D, levels have also been detected in the synovial fluid of joints in rheumatoid arthritis sufferers. It has been suggested that, as with the lung, SP-D may be involved in phospholipid homeostasis within the joint as well as acting as a dual pro- and anti-inflammatory mediator (Kankavi, 2006). Reduced levels of SP-D in serum and raised levels in synovium may contribute to the persistent low-grade joint inflammation seen in rheumatoid arthritis patients (Brown *et al.,* 2006).

SP-D synthesis has been demonstrated in localised vascular endothelial cells in both mice and humans. SP-D-null mice fed an atherogenic diet showed significant protection against atherosclerotic legions in the aorta, and a small foamy macrophage count compared to control mice. Nevertheless, foamy macrophage masses were found in the lung - a typical feature in SP-D deficiency. When treated with a recombinant SP-D fragment anti-inflammatory Th1 T-helper cells were polarised to allergic Th2 phenotypes, and an upregulation of pro-inflammatory TNFα was observed. Also seen was a reduction of plasma lipid concentrations: High-density lipoprotein levels were significantly elevated, while low-density cholesterol was decreased. Overall, SP-D is viewed as pro-atherogenic in the mouse model used - probably due to disturbances of plasma lipid metabolism and alteration of the inflammatory process (Sorensen *et al.,* 2006).

Skin diseases such as psoriasis and atopic dermatitis have recently been linked to changes in SP-D levels in tissues. When immunohistochemically stained, affected skin biopsies showed no difference in SP-D reactivity compared to normal skin. However, in the stratum spinosum (the middle layer of the epidermis where keritinisation is initialised), SP-D showed higher immunostaining intensity. Serum SP-D levels were unchanged, suggesting that - unlike pulmonary disorders - there is no spill-over of SP-D from affected lesions. The same study also reported that production of SP-D transcripts was not upregulated, and the increase in infiltrating inflammatory cells may be the source of the protein in the lesions (Hohwy *et al.,* 2006).

1.2.7 Pregnancy and development

Pulmonary surfactant is secreted from the foetal lung into the amniotic fluid where it accumulates during the third trimester of pregnancy (Bayer *et al.,* 1973). Respiratory failure is a major cause of neonatal death in preterm babies, especially those born before 32 weeks and survival is closely associated with development of surfactant. During labour, intrauterine tissue macrophages release prosataglandin, along with other macrophage-associated cytokines such as pro-

inflammatory TNFα. Surfactant provides the amniotic epithelium with a source of constituents used for prostaglandin synthesis during labour.

SP-D can be detected in human amniotic fluid as early as 26 weeks' gestation, and has been found in foetal membranes of the amniotic epithelium and chorionic membrane as well as certain layers of the late pregnant uterus (Miyamura *et al.,* 1994). It is likely that the presence of SP-D is necessary for clearance of pathogens and activation of macrophages - especially with regards to intrauterine infections. SP-D serves as a well-suited, non-specific immune molecule in the uterus, as well as maintaining integrity of the foetal membranes (Hills, 1994).

The biological role of SP-D in pregnancy is not fully understood, but involvement in tissue remodelling has been suggested, due to an interaction with decorin - a proteoglycan that binds collagen and other extracellular matrix components (Nadesalingam *et al.,* 2003). SP-D and decorin interact by either the CRD of SP-D binding to the carbohydrate moieties of decorin, or by decorin binding to SP-D's collagen region. Decorin induces expression of matrix metalloproteases and disrupts collagen fibres to facilitate cervical dilation during labour. Co-purification of SP-D with decorin in amniotic fluid suggests a possible role in intrauterine tissue remodelling during parturition (Kishore *et al*., 2005).

Extensive localisation of SP-D in the female reproductive tract suggests a crucial role in local host defence against microorganisms, especially during the secretory phase of the menstrual cycle when it is necessary to protect the uterus during implantation of the embryo. SP-D has been shown to enhance phagocytosis of chlamydial pathogens, and inhibit *Chlamydia trachomatis* infection of cervical epithelial cells (Oberley *et al.,* 2004).

1.2.8 Immune-related functions of SP-D

In addition to phagocytosis, SP-D regulates multiple cellular responses by controlling inflammatory function. As mentioned previously, SP-D shows both pro- and anti-inflammatory signalling properties and recent focus on the effect of S-nitrosylation of N-terminal cysteines has shed light on exactly how SP-D is able to regulate inflammation (see Fig 1-16).

Nitric oxide (NO) is required for control of vessel and airway dilation in the lung and is a key mediator of acute lung injury, bronchopulmonary dysplasia, respiratory distress syndrome and asthma (Ricciardolo *et al*., 2004). The role of NO in signal transduction pathways has gained significant attention in recent years, especially in regards to its capability to induce signalling functions via post-translational modification of proteins (Stamlet *et al*., 2001). S-nitrosylation, the covalent modification of thiols by formation of S-nitrosothiol (SNO) groups, has emerged as an important regulator of protein function. As mentioned previously, oligomerisation of SP-D is dependent on cysteine residues at positions 15 and 20 in the N-terminal region - and mutations of these results in SP-D being unable to form quaternary structures larger than single trimers (Brown-Augsburger *et al*., 1996). It was proposed that these residues were likely targets for NO activity, and as such may alter the oligomer structure and regulate inflammation. Subsequent studies by Guo *et al.,* 2008 demonstrated that nitrosylation results in a functional 'switch' that activates the pro-inflammatory ability of SP-D.

Their results demonstrated that SP-D can be nitrosylated both *in vivo* and *in vitro* and that this is associated with disruption of the multimeric structure. Multimers in the lung are able to interact with invading pathogens and with cell surface receptor signal-inhibitory regulatory protein (SIRP) α. Binding to the latter causes a suppression of p38 phosphorylation by SHP-1, which in turn reduces activation of NF-κβ. Conversely, it is presumed that interactions of the CRD with pathogenic molecules results in more trimers of SP-D being 'pulled' from the multimer, exposing their tail domains - which are then able to interact with calreticulin/CD91. Binding of this

complex rather than SIRPα causes the opposite effect on p38, which is phosphorylated to activate NF-κβ and its downstream inflammatory responses.

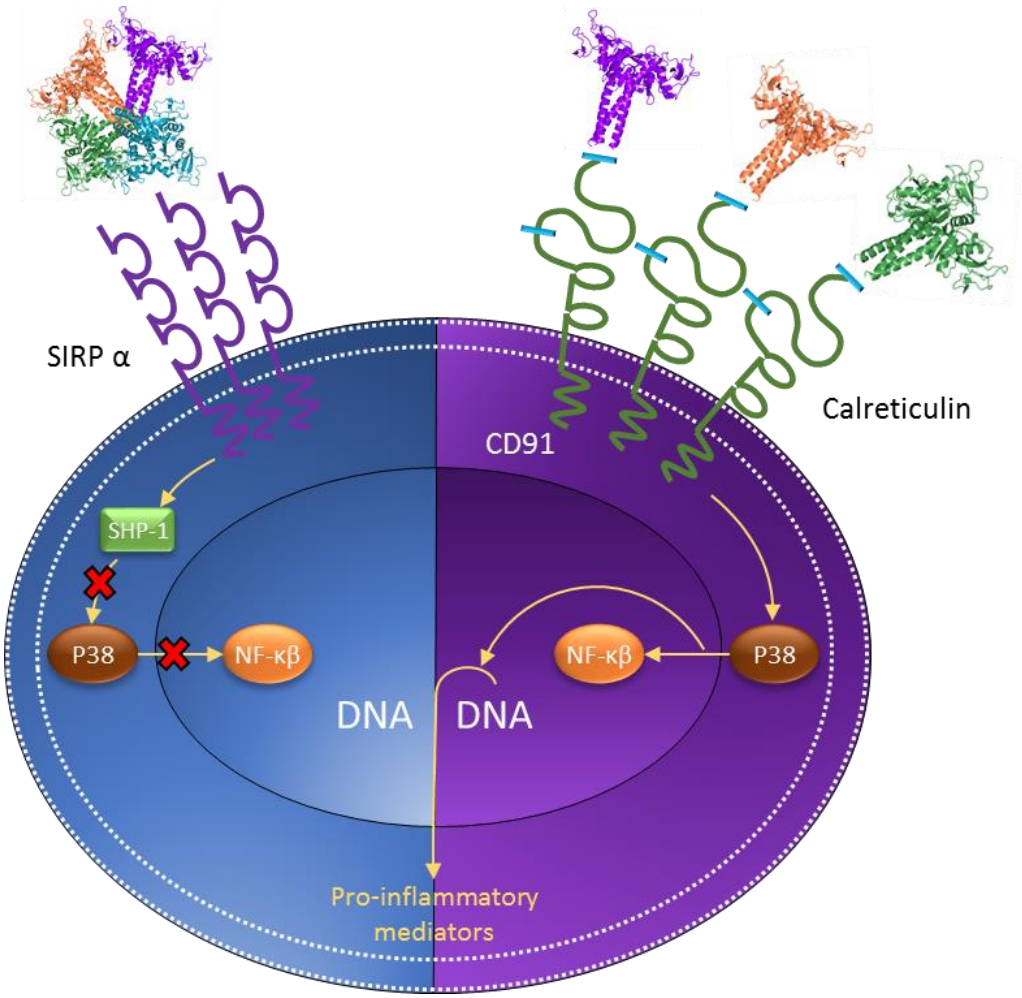


Figure 1-16 The pro- and anti-inflammatory functions of SP-D. Under non-inflammatory conditions SP-D remains in a dodecameric forms with the tail domains buried, CRDs bind to SIRPα and activates SHP-1 which in turn inhibits p38 and NF-κβ activation. Under inflammatory condition SP-D multimeric structure is disrupted and binds to calreticulin, p38 is phosphorylated via CD91 and NF-κβ is stimulated to induce a response. Image edited from *Guo et al.,* 2008.

1.3 Characterisation of SP-D

Meticulous characterisation of a recombinant head and neck fragment by Shrive *et al*., 2003, detailed the high-resolution structures of both native and ligand-bound human SP-D. Both the native and maltose-bound structures contain residues 205-355 in chains A and B, and 206-355 in

chain C. The overall structures reveal a trimeric assembly of an extended α-helical coiled-coil neck (residues 203-235) which transitions at a *cis*-proline at position 235 into three globular CRD domains (residues 236-355). The N-terminal collagen region (the first 24 residues, 179-202) was not visible in the electron density maps, along with the first few residues of the neck-region. Three calcium ions per CRD are clearly visible in the map, with an additional fourth ion located in the central pore formed by the three subunits, at the neck-CRD interface.

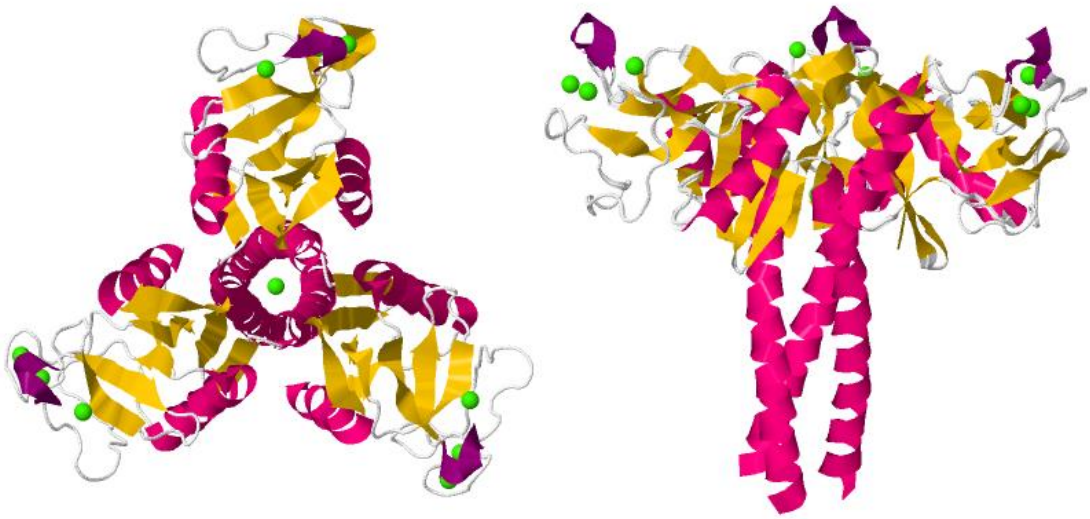


Figure 1-17 Crystal structure of the native recombinant head and neck fragment of human SP-D showing the general trimeric assembly and three calcium ions per CRD (green spheres), as well as the fourth calcium ion in the central pore. The left-hand image shows the structure viewed down the molecular 3-fold axis, the right-hand image show the perpendicular view. Images generated using JMol, coordinates taken from PDB code 1PW9.

1.3.1 Calcium ions in the CRD

The distances of coordinating bonds of the ligand-binding calcium ion (here on referred to as Ca1) vary across the three chains, and involve the sidechains of Glu321, Asn323, Glu329, Asn341, Asp342 and their carbonyl group. Two hydroxyl groups from the incoming carbohydrate ring complete coordination; in the absence of bound ligand coordination is completed by two water molecules (see Fig 1-18). Native structures see bonding distances between 2.23 and 2.76 Å, and

between 2.28 and 2.70 Å for maltose-bound SP-D, with no overall difference visible between native and ligand-bound protein structures.

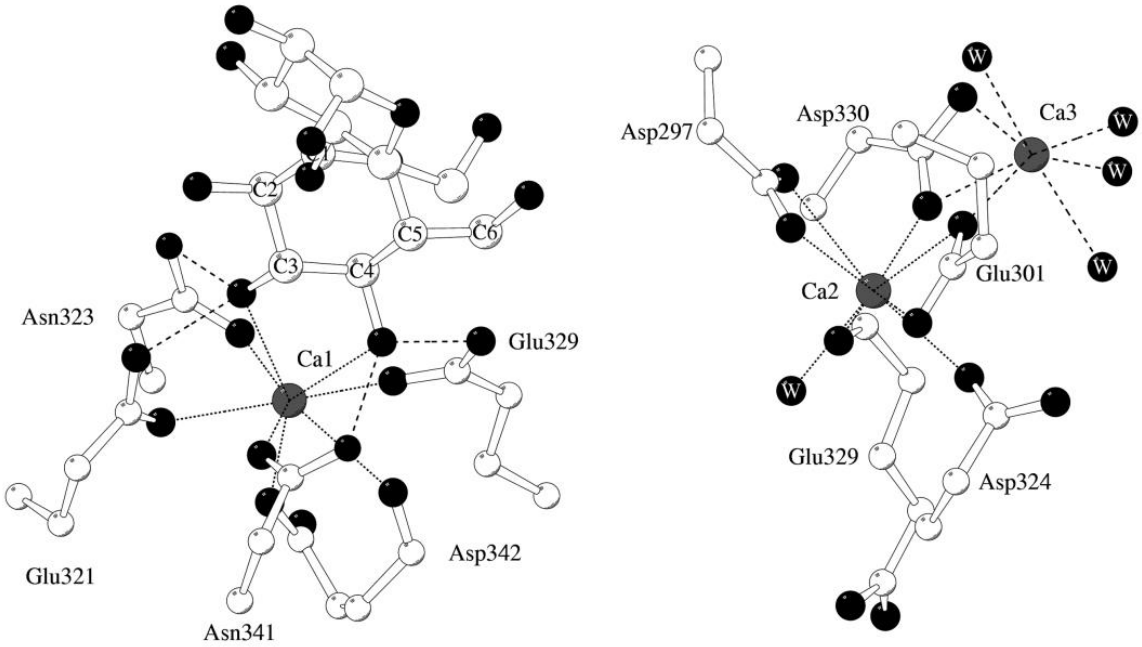


Figure 1-18 Coordination of calcium sites in SP-D. The left-hand image depicts coordination of Ca1, as well as interaction of maltose via two vicinal hydroxyl groups on the pyranose, in the absence of ligand coordination is completed by two water molecules. The right-hand image shows coordination of calcium ions Ca2 and Ca3. Image adapted from Shrive *et al.,* 2003.

In addition to Ca1, the CRDs of both ligand-bound and native structures contain two additional ions, Ca2 and Ca3 in each chain. Located on the surface of the CRD, they lie in close proximity to Ca1, separated by 3.9-4.0 Å. Ca2 is coordinated by the acidic sidechains of Asp297, Glu301, Asp324, Asp330, a water molecule and the main-chain carbonyl group of Glu329 whose sidechain also participates in coordination of Ca1; while Ca3 is coordinated by the acidic sidechains of Glu301 and Asp330, as well as four water molecules (see Fig 1-18). The fourth calcium, seen in the native SP-D structure only, is coordinated by Glu232 of all three chains and three water molecules.

The following table details the interactions of calcium with amino acid sidechains and water molecules/maltose hydroxyl groups, along with their distances.

Table 1-10 Coordination of calcium by amino acid sidechains and water molecules in the CRD domain. Distances are for native SP-D (with exception of Ca1 coordination by ligand hydroxyl groups) and show the range in chains A, B and C. Values taken from Shrive *et al.,* 2003.

Calcium atom	Amino acid sidechain atom	Distance in Å
Ca1	Glu321OE1	2.53-2.73
	Asn323OD1	2.37-2.52
	Glu329OE1	2.30-2.41
	Asn321OD1	2.36-2.37
	Asp342OD1	2.23-2.34
	Asp342O	2.68-2.76
	Ligand/water	2.51-2.58
	Ligand/water	2.54-2.68
Ca2	Asp297OD1	2.59-276
	Asp297OD2	2.42-2.52
	Glu301OE1	2.57-2.60
	Glu301OE2	2.50-2.68
	Asp324OD1	2.47-2.58
	Glu329O	2.38-2.49
	Asp330OD1	2.49-2.59
	Water	2.23-2.36
Ca3	Glu301OE1	2.36-2.57
	Asp330OD1	2.46-2.58
	Asp330OD2	2.50-2.57
	Water	2.16-2.23
	Water	2.30-2.57
	Water	2.25-2.45
	Water	2.25-2.49
Ca4	Glu232OE1	2.21-2.38
	Water	2.22-2.45

There is a marked asymmetry in the conformation of tyrosines at position 228 in the neck region due to depletion of calcium 4. The sidechain of Tyr228 in the C chain is shifted into the centre of the coiled-coil, with the remaining two directed outward. Although the conformation of

the tyrosine is fundamentally identical in both native and maltose-bound structures, the interactions between it and Glu232 in the neck, and Lys246 in the CRD, are significantly different (see Fig 1-19).

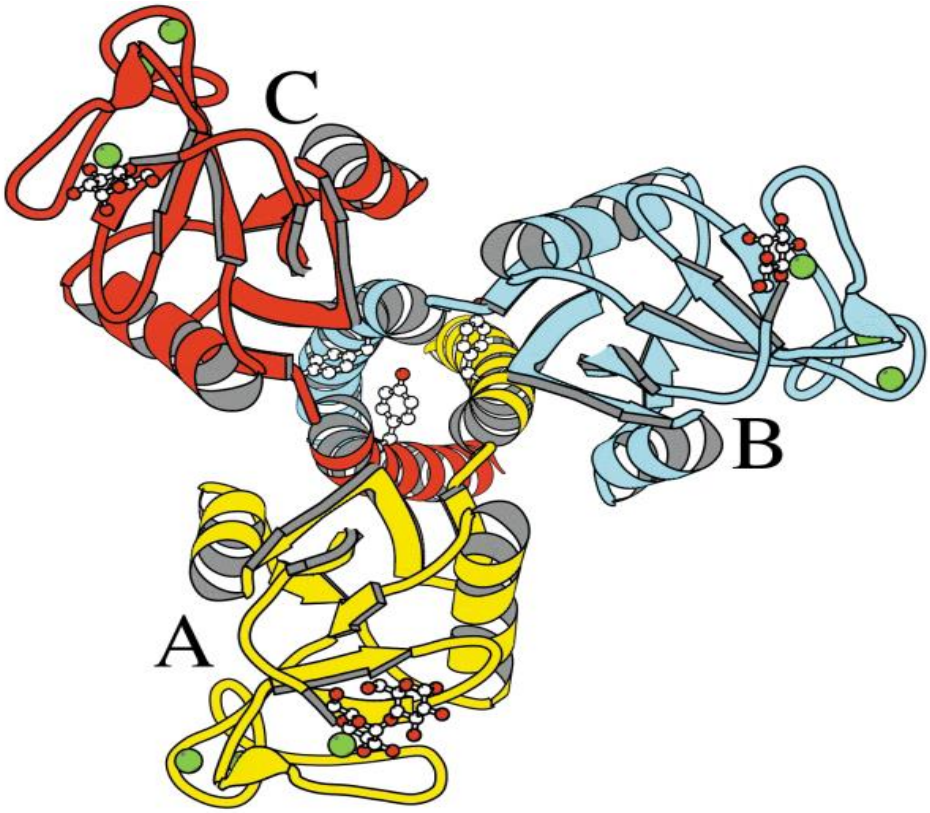


Figure 1-19 Crystal structure of maltose-bound SP-D showing three calcium ions per CRD in green, and the asymmetric tyrosine positioned into the central pore. Image taken from Shrive *et al.,* 2003.

In the absence of Ca4, two of the three glutamate residues at 232 interact with each other as well as with Lsy246 in the same chain, while the third (in the B chain) adopts a conformation whereby it interacts with the asymmetric tyrosine 228. In the structure without bound ligand (and with calcium 4 present) all three glutamate residues assume similar conformations and interact with both Lys246 in the same chain, and with each other.

The electron density of native SP-D suggests that the asymmetric Tyr228 interacts with the main-chain carbonyl of Tyr228 in the B chain through a hydrogen-bond bridge with a water molecule. This water molecule may occupy the position otherwise taken by the shifted Glu232; however, it is more likely an indication of two conformations of Glu232 (Shrive *et al.,* 2003). The resulting asymmetry of Tyr228 and the interactions of Glu232 do not disrupt the geometry of the neck-CRD relationship.

1.3.2 Other crystal structures of ligand-bound SP-D

Surfactant protein D has been shown to bind a range of monosaccharides via two vicinal hydroxyl groups on the pyranose ring or by the glycerol sidechain in heptose. Binding mechanisms can be divided into subgroups depending on the vicinal hydroxyl groups that SP-D binds:

- Mannose-type via O3 and O4 hydroxyl groups on the pyranose ring (including glucose, maltose, maltotriose, p-nitrophenyl maltoside, N-acetyl mannosamine, α-methyl mannoside, trimannose, α1-2mannobiose and α1-4mannobiose)
- galactose via O1 and O2
- myoinositol via O1 and O6
- fucose via O2 and O3
- phospholipid type binding/inositol-phosphate via O4 and O5
- heptose via the glycerol tail groups O6 and O7.

Despite this, the mechanism of binding is similar with regards to the spatial positioning of the interacting hydroxyl groups of the carbohydrate, and involves a conserved group of amino acids. An example of hydroxyl-pair positioning is given below to illustrate this similarity.

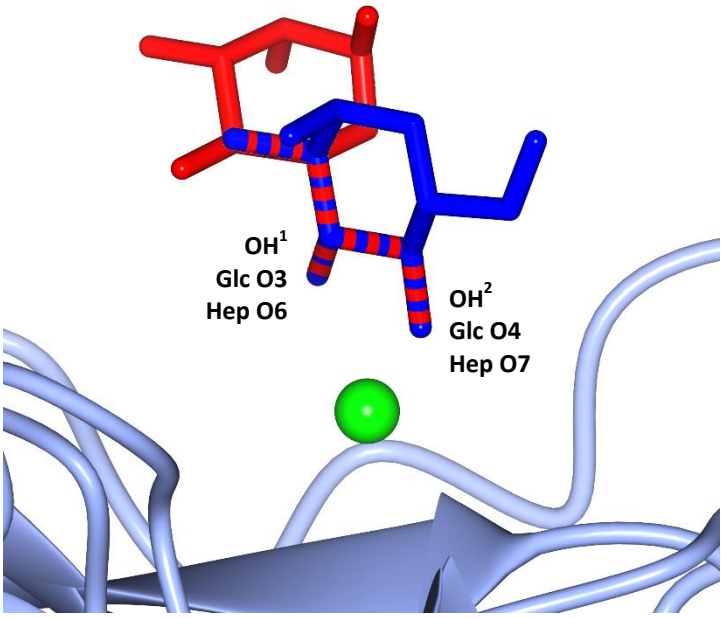


Figure 1-20 Binding of glucose via vicinal hydroxyl groups O3 and O4 (Glc, blue) and heptose via hydroxyl groups of the glycerol side chain O6 and O7 (Hep, red). Bonds shown in alternate colours share the same coordinates in binding. OH^1 can be the O2 of galactose, O3 of α1-2mannobiose and α1-4mannobiose (in chain A only), O4 of α1-4mannobiose (in chains B and C), O4 of inositol-1-P (in chains B and C) and the O3 of maltose. OH^2 can be the O1 of galactose, O4 of α1-2mannobiose and α1-4mannobiose (in chain A only), O3 of α1-4mannobiose (in chains B and C), O5 of inositol-1-P (in chains B and C) and the O4 of maltose. Binding of myoinsoitol utilises the O1 and O6 hydroxyl groups which can take positions OH^1 and OH^2 reversibly. Image generated using CCP4MG using PDB accession codes 1PWB and 2RIB.

Numerous ligand-bound SP-D structures have been characterised, comprising mainly of monosaccharides but also longer chain ligands such as di- and trisaccharides.

Table 1-11 Ligand-bound SP-D structures deposited in the Protein Data Bank in chronological order, their spacegroup and hydroxyl groups of the terminal saccharide involved in binding.

Ligand	Spacegroup	Mechanism of SP-D binding	Accession code and reference
Maltose	$P2_1$	Via O3 and O4	1PWB, Shrive *et al.,* 2003
Maltotriose	$P2_1$	Via O3 and O4	2GGU, Crouch *et al.,* 2006

Table 1-12 Ligand-bound SP-D structures deposited in the Protein Data Bank in chronological order, their spacegroup and hydroxyl groups of the terminal saccharide involved in binding.

p-nitrophenyl maltoside	P21	Via O3 and O4	2GGX, Crouch *et al.,* 2006
N-acetyl mannosamine	$P2_1$	Via O3 and O4	2ORJ, Crouch *et al.,* 2007
Inositol-1-phosphate	$P2_1$	Via O4 and O5	2ORK, Crouch *et al.,* 2007
Myoinositol	$P2_1$	Via O1 and O6 (reversible)	2OS9, Crouch *et al.,* 2007
D-glycero-D-manno-heptose	$P2_1$	Via O3 and O4	2RIA, Wang *et al.,* 2008
L-glycero-D-manno-heptose	$P2_1$	Via O6 and O7	2RIB, Wang *et al.,* 2008
L-glycero-D-manno-heptopyranosyl-(1-3)-L-glycero-D-manno-heptopyranose	$P2_1$	Via O6 and O7	2RIC, Wang *et al.,* 2008
Allyl 7-O-carbamoyl-L-glycero-D-manno-heptopyranoside	$P2_1$	Via O3 and O4	2RID, Wang *et al.,* 2008
2-deoxy-L-glycero-D-manno-heptose	$P2_1$	Via O6 and O7	2RIE, Wang *et al.,* 2008
α-methyl mannoside	$P2_1$	Via O3 and O4	3G81, Crouch *et al.,* 2009
α1-2 mannobiose	$P2_1$	Via O3 and O4	3G83, Crouch *et al.,* 2009
α1-4 mannobiose	$P2_1$	Via O3 and O4	3G84, Crouch *et al.,* 2009
Galactose	$P2_1$	Via O1 and O2	3IKN, Shrive *et al.,* 2009
Inositol-1-phosphate	$P2_1$	Via O4 and O5	3IKP, Shrive *et al.,* 2009
α1-2 mannobiose	$P2_1$	Via O3 and O4	3IKQ, Shrive *et al.,* 2009
α1-4 mannobiose	$P2_1$	Via O3 and O4	3IKR, Shrive *et al.,* 2009
Haemophilus Eagan 4A polysaccharide	$P2_1$	Via O6 and O7 of heptose	4E52, Clark *et al.,* unpublished

Four residues that coordinate the calcium Ca1 also interact with the carbohydrate: the OE2 of Glu321 and ND2 of Asn323 interact with the OH^1 hydroxyl group, and the OE2 of Glu329 and ND2 atom of Asn341 interact with the OH^2 group (see Fig 1-18 and 1-20).

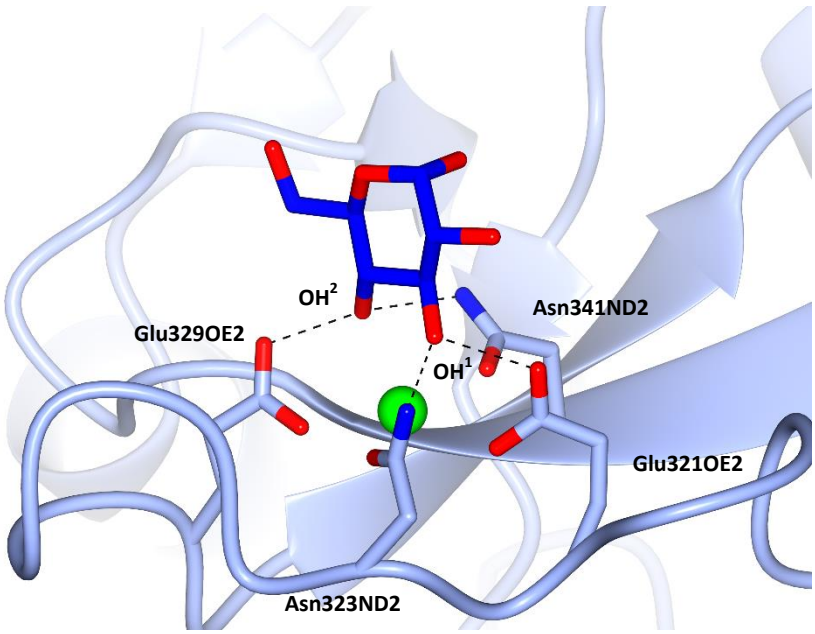


Figure 1-21 Glucose coordination by the amino acid sidechains of Glu321, Asn323, Glu329 and Asn341 in the C chain. Protein shown in light-blue, glucose in dark blue and calcium ion Ca1 shown as the green sphere. Image generated using CCP4MG, from coordinate file 1PWB.

As mentioned previously, in a protein crystal contacts occur between atoms of macromolecules. They are as such, an artefact of crystallisation, but may influence the binding of ligands if in close proximity to the CRD. The amino acid Asp325 which flanks the binding site, has been directly implicated in influencing crystal contacts by Shrive *et al*., 2009; in all three chains it forms bonds with neighbouring trimers via a contact region spanning residues 322-326. Chains B and C form similar, but non-identical contacts: both chains include charge-charge interactions between Asp325 and Lys230 of the neighbour, hydrogen bonds between the carbonyl group of Pro322 and Gln222, and between Asn323 and Ser226. The latter contact formed by Asn323, which coordinates bound ligand, causes a rotation of this sidechain, moving the ND2 atom towards serine and away from the ligand-binding pocket compared to chain A. The result is a small but significant shift in ligand orientation (see Fig 1-22).

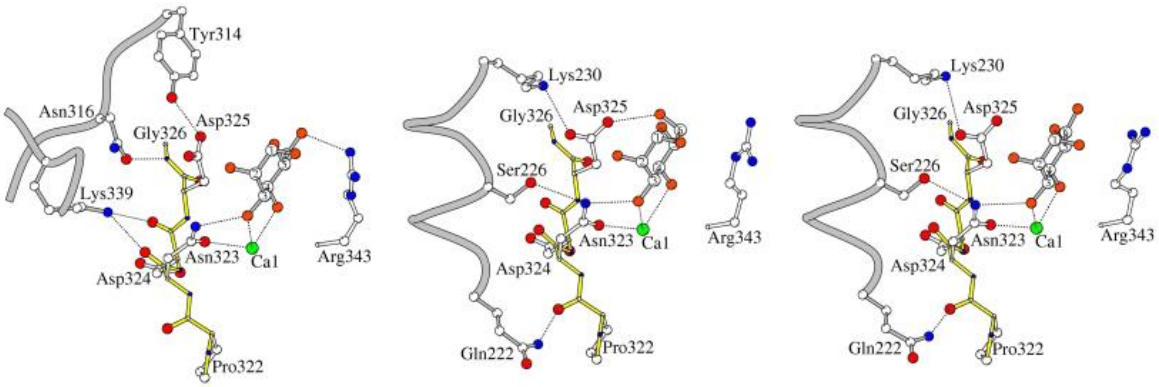


Figure 1-22 Influence of crystal contacts on bound α1-2mannobiose in chain A (left), B (centre) and C (right). The primary protein main-chain is shown in yellow, coordinating residues involved in ligand-binding are shown in white and the neighbouring crystal contact protein in grey. Figure taken from Shrive *et al*., 2009.

The crystal contact between the A chain and neighbouring protein is considerably different to that seen in the B and C chains as the neighbouring molecule sits in closer proximity to the CRD, resulting in a more restricted ligand binding site. Contacts include bonds between the main chain carbonyl of Asp324 (whose sidechain coordinates Ca2) and the sidechain of ligand-binding Asn323 with Lys339 of the neighbouring molecule, and the main chain N of Gly326 with Asn316. In addition, Asp325 forms a hydrogen bond with Tyr314 of chain B of the neighbour and the sidechain adopts a different conformation to that seen in the B or C chain.

The position of Asp325 affects the position and coordination of bound ligand, which in turn influences the sidechain conformation of Arg343 on the opposite side of the binding pocket. As such, the orientation of Arg343 varies between native and maltose-bound structures. The 'penultimate' glucose saccharide that does not interact with Ca1 is only visible in the electron density in this chain, as it forms a stabilising hydrogen-bond bridge with the main chain carbonyl group of Ala275 of chain B of the symmetry-related trimer. This contact results in an approximate 10° tilting of the bound, terminal glucose toward Arg343 rather than Asp325. The altered configuration of both glucoses is such that they interact through a 2.7 Å hydrogen-bond (Shrive *et al.,* 2003). This mechanism is not seen in either mannobiose structures, but orientation of Arg343

differs between various ligand-bound structures (Shrive *et al.,* 2009). It has been suggested that Arg343 plays a role in discrimination between ligands by influencing entrance to the binding pocket (Allen, 2004), but in some cases multiple ligand orientations are observed. It is more likely that the crystal environment itself is responsible; variations in side-chain and/or ligand orientation illustrate flexibility of the protein, and how SP-D is able to make minor adjustments to bind different patterns of surface carbohydrates on pathogen surfaces.

SP-D has also been shown to bind longer ligands such as maltotriose and *p*-nitrophenyl maltoside whereby in addition to the terminal carbohydrate binding, phenylalanine 335 outside the binding site interacts with the third glucose saccharide. Site-directed mutagenesis demonstrated that this residue (or equivalent aromatic amino acid at this position) is necessary for 'stacking' of the ligand over it (Crouch *et al.,* 2006).

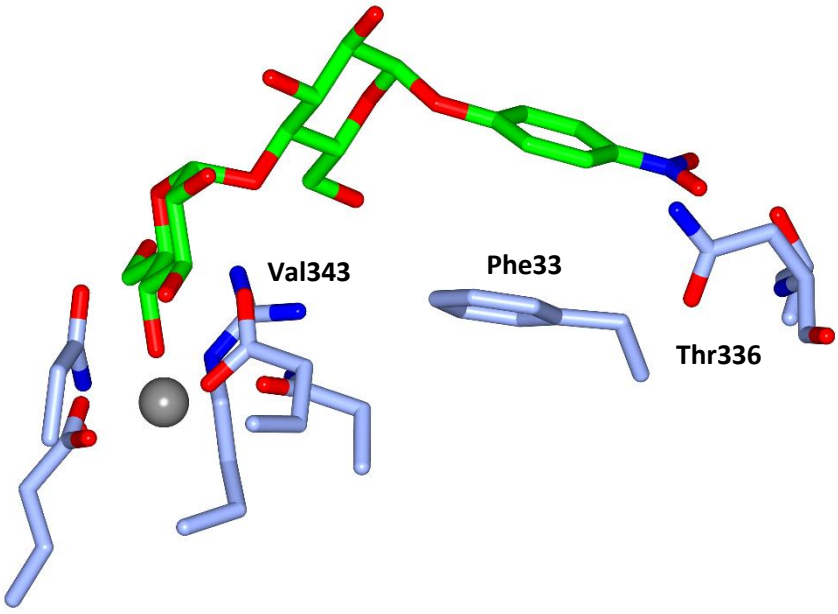


Figure 1-23 Crystal structure of *p*-nitrophenyl maltoside (green) binding by SP-D; showing the ligand coordination by Val343, Phe335 and Thr336. Coordinating SP-D residues are shown in blue, with calcium as the grey sphere. Image generated using CCP4MG using PDB accession code 2GGX

Modelling studies have demonstrated that binding via the principal CRD calcium ion can also occur through non-terminal saccharides. Unlike other structures, trimannose (Man-α1,2-Man-α1,2-Man) was crystallised in the $P2_12_12_1$ spacegroup due to site-directed mutagenesis of Ala

to Asp at position 325, and Val to Arg at 343. As Asp325 and Arg343 are directly implicated in forming crystal contacts, the spatial arrangement of symmetry-related protein molecules results in a more spacious binding pocket - and an elongation of the long loop in the CRD that increases flexibility. Binding was essentially similar to that seen of maltose, and coordination occurred via the O3 and O4 hydroxyl groups. However binding was observed for the non-reducing saccharide in the chain, rather than through either terminal mannose (Goh *et al.,* 2013).

In addition, a recently elucidated structure of bound polysaccharide illustrates SP-D's ability to bind non-terminal monosaccharides in longer chain trisaccharide ligands. The polysaccharide, consisting of glucose, L-glycero-D-manno-heptose and 3-Deoxy-D-manno-oct-2-ulosonic acid (Kdo) is bound via the O6 and O7 groups of the non-reducing heptose (Clark *et al*., unpublished).

1.4 Pathogen binding by SP-D

The following subchapter details the structural evidence for the binding mechanism of SP-D to pathogens known to be neutralised by SP-D.

1.4.1 Bacterial recognition

Surfactant protein D has been shown to bind predominantly to heptose residues in the inner core of bacterial lipopolysaccharides. Studies by Wang *et al.*, 2008 demonstrated that binding to heptose occurs via the O6 and O7 hydroxyl groups, and presented models of both monosaccharide and disaccharide heptose binding (see Fig 1-24). Interactions with the glycerol sidechain provide a potential mechanism of bacterial recognition that does not involve hydroxyl groups on the pyranose ring - and because the core structure of LPS is highly conserved, this suggests a common binding mode of SP-D to a range of Gram-negative bacteria.

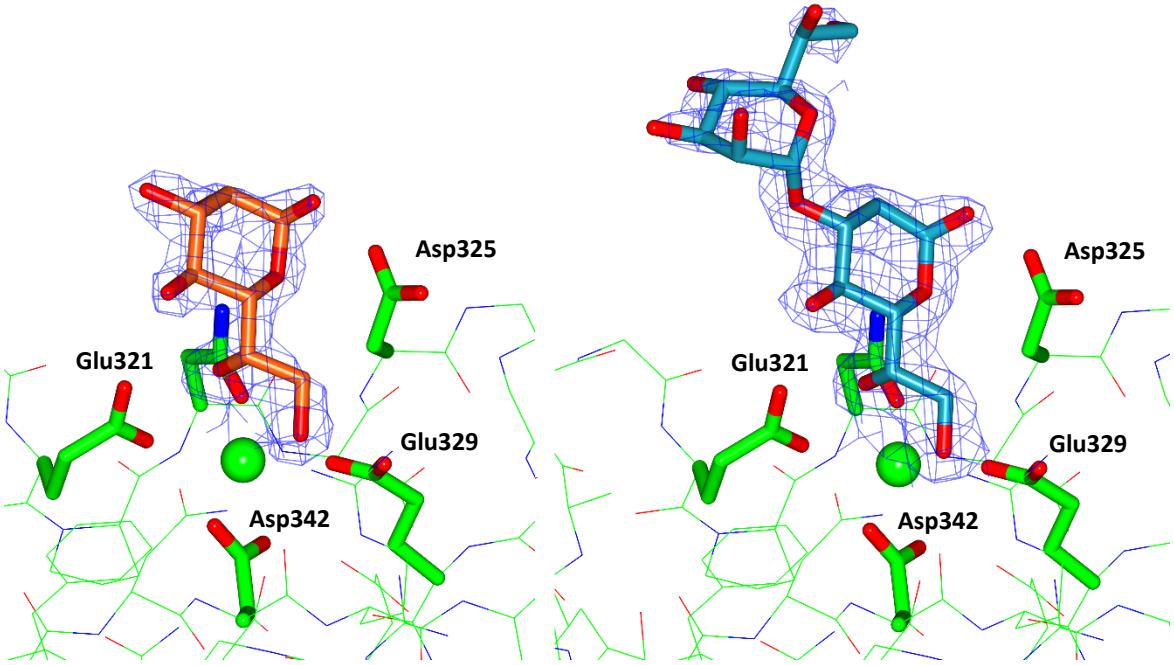


Figure 1-24 Crystallographic complex of SP-D with heptose monosaccharide (left, orange) and diheptose (right, cyan). The diheptose is linked via an α1-3 bond, corresponding to the first two heptose residues in LPS. Ligand-coordinating CRD residues are green and red, with calcium ions shown as green spheres and electron density in blue. Image generated using CCP4MG using PDB accession codes 2RIB and 2RIC.

In addition to recognition of the inner core structure of LPS, SP-D has been shown to bind some bacterial species via the O-antigen. *Helicobacter pylori* and nonencapsulated *Klebsiella* strains (with O-antigen repeating units that contain fucose and mannose residues with accessible hydroxyl pairs, respectively) are bound and agglutinated by SP-D and initiate the same innate immune response as strains that bind by the inner core (Khamri *et al.,* 2005; Sahly *et al.,* 2002). No evidence has yet demonstrated SP-D's ability to directly bind H-antigens, however recent studies have shown that SP-D binds to toll-like receptors that do (Wright, 2005).

Strains of *Klebsiella pneumoniae* that express mannose-rich capsular polysaccharides have been shown to be recognised by certain C-type lectins, which results in phagocytosis of the cell (Kostina *et al.,* 2005), but studies with SP-D have yet to be undertaken. Furthermore, studies reveal that SP-D is able to bind PepG of *S. aureus* in a calcium-dependant manner, strongly

suggesting that the binding occurs via the CRD (de Wetering *et al.,* 2001). Binding of Gram-negative bacteria by SP-D results in agglutination, upregulated phagocytosis by macrophages and inhibited growth (Kuan *et al*., 1992; Wu *et al*., 2003).

SP-D is able to bind lipoteichoic acid moieties of Gram-positive bacterium *B. subtilis* in a calcium-dependant manner, strongly suggesting that binding occurs via the CRD (de Wetering *et al.,* 2001). When binding profiles of LTA and PepG were compared it was observed that higher concentrations of the latter were required to block binding to LTA, suggesting that SP-D has a higher affinity for LTA than to PepG. This may reflect the greater importance of lipoteichoic acid in lung infection as it is shed during growth and plays a role in attachment of the bacteria to host epithelial cells (Sutcliffe & Shaw, 1991). Although aggregation occurred, no enhanced killing of Gram-positive *Streptococcus pneumoniae* was seen (Jounblat *et al.,* 2004).

1.4.2 Mycobacterial recognition

SP-D binds lipoarabinomannan (LAM) present on both *M. tuberculosis* and *M. avium* and reduces uptake of the former - reducing adherence of bacteria to macrophages while inducing agglutination, but enhances phagocytosis of the latter. It has been suggested that these opposing effects are due to slight structural differences between LAM molecules on bacterial cells. The arabino terminus of LAM derived from *M. avium* is capped with a single mannose residue, rather than the more common dimannose (Ferguson *et al.,* 1999; Kudo *et al.,* 2004). The enhanced phagocytosis of *M. avium* is believed to be as a result of upregulation of cell surface receptors on macrophages, achieved without the need for direct binding (Beharka *et al.,* 2002).

1.4.3 Viral recognition

SP-D is known to bind a number of viruses resulting in their neutralisation and upregulated phagocytosis. SP-D has been found to bind mannosylated, N-linked carbohydrates on the Influenza A virus haemagglutinin and neuraminidase domains via the CRD (Hartshorn *et al.,* 2000).

Sensitivity of IAV recognition by SP-D is dependent on glycosylation at specific sites, which must be in close proximity (<20 Å) to the sialic binding site (Hartshorn *et al*., 2008).

Studies by Meschi and coworkers in 2005 demonstrated that SP-D also binds the envelope protein gp120 of human immunodeficiency virus (HIV), containing a highly conserved mannosylated oligosaccharide (see Fig 1-14). Glycoprotein gp120 contributes to the antibody resistance of HIV and evades binding of cell surfaces of macrophages and dendritic cells, but SP-D was shown to bind gp120 and inhibit HIV infectivity. Binding is mediated by calcium-dependant carbohydrate binding, and like that of IAV is governed by the glycosylation extent of the viral protein.

SP-D has also been shown to bind bovine strains of non-enveloped rotavirus via VP7 glycoproteins on the viral surface, resulting in a range of antiviral activities including agglutination, enhancement of interactions with neutrophils and neutralisation of infectivity. Glycans of VP7 are all of the high-mannose type (Reading *et al*., 1998).

SP-D binds to RSV via viral envelope glycoproteins G and F *in vitro* and *in vivo* in a calcium dependent-manner, inhibiting viral growth and reducing levels of the virus in the lung by 80% (Hickling *et al.,* 1999). G proteins contain O- and N-linked oligosaccharide chains, typically high-mannose N-linked polysaccharides (Wertz & Lichtenstein, 1988). F-protein N-glycosylation was analysed and illustrated that the carbohydrate moieties are complex-oligosacchrides, comprised of fucose, glucosamine and galactosamine, with a mannose core (Rixon *et al,* 2002). The exact mechanism of SP-D recognition has yet to be elucidated.

SP-D serum levels have been shown to rise during infection with the SARS (severe acute respiratory syndrome) coronavirus. Viral particles contain nucleocaspid proteins that interrupt normal host cell cycles to reduce interferon production. SP-D has been shown to strongly inhibit coronaviral activity by binding directly to ‘spike’ (S) glycoproteins and induce agglutination (Wu *et al.,* 2009). S proteins contains a single N-linked glycosylation site that facilitates this inhibition,

and although not fully elucidated, it is believed to be a high-mannose oligosaccharide (Zhou *et al*., 2010).

The recognition of high-mannose oligosaccharides by SP-D is yet to be fully understood, but in 2011 Goh and coworkers modelled a possible binding mechanism through a large-scale molecular dynamics simulation. They hypothesised that the Man-α1,2-Man-α 1,2Man trisaccharides were the likely target of binding, and using the known conformation of the two terminal residues by Shrive *et al.,* 2003, *superimposed* the remaining saccharides (see Fig 1-25). They concluded that despite the flexibility of glycosidic linkages, high mannose glycans such as those found on viral surfaces have a limited number of possible conformations. Sensitivity of IAV recognition by SP-D is dependent on glycosylation at specific sites, which must be in close proximity (<20 Å) to the sialic binding site (Hartshorn *et al*., 2008).

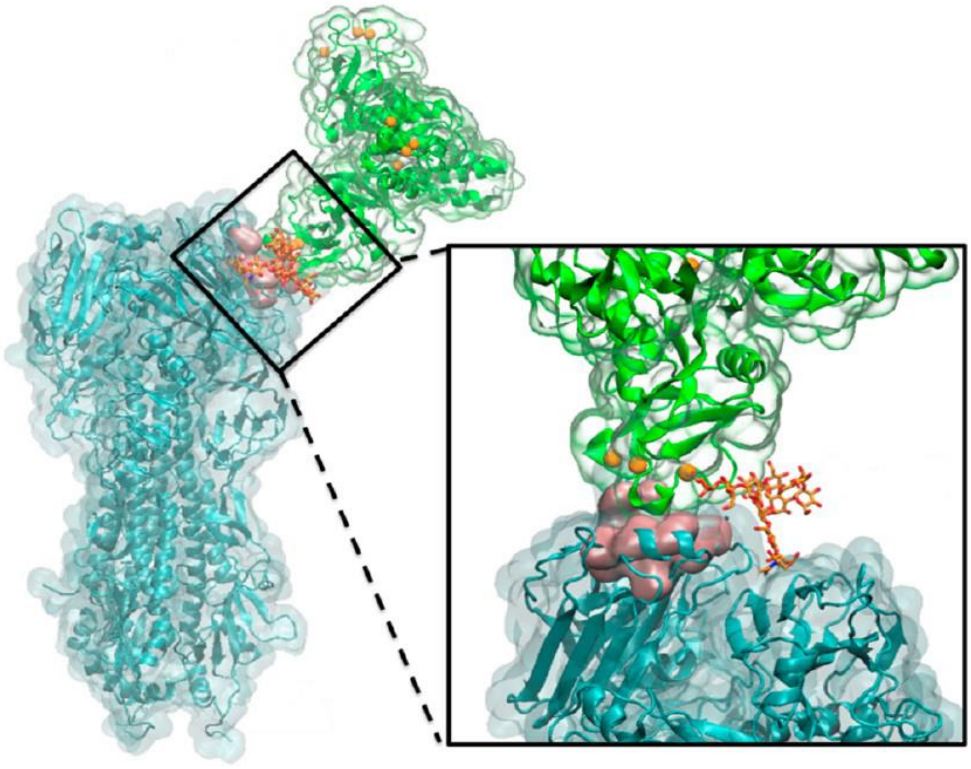


Figure 1-25 Simulated structure of a high mannose oligosaccharide (orange) in proximity to the sialic acid binding site of haemagglutinin (pink). The remainder of the HA molecule shown in cyan and SP-D is shown in green. The orange spheres represent calcium ions of the CRD. Taken from Goh *et al.,* 2011.

1.4.4 Fungal recognition

Phagocytosis and killing of *Aspergillus fumigatus* and *Cryptococcus neoformans* by alveolar macrophages and neutrophils has been observed as a result of SP-D binding. *A. fumigatus* is bound by SP-D calcium-dependently, through N-linked oligosaccharides on glycoproteins 45 and 55 (Madan *et al.,* 1997). *C. neoformans* and *Histoplasma capsulatum* are bound by capsular polysaccharide glucuronoxylomannan (consisting of mannose, xylose and glucuronic acid), and mannoprotein-1 - a protein of about 100 kDa attached to β1-6-glucan through a glycosylphosphatidylinositol anchor, containing five alpha-linked mannosyl residues (Richardson, 1993; Kollar *et al.,* 1997).

Blastomyces dermatitidis and *Pneumocystis jirovecii* are bound specifically by β1-6-glucan, resulting in aggregation and increased attachment to macrophages. Glucans containing β1-3 links are not bound by SP-D, as the otherwise occupied O3 hydroxyl group is required for calcium-dependent recognition (Allen *et al.,* 2001b).

SP-D has been shown to bind to *Candida albicans* and directly inhibit growth by aggregation of the organism without involvement of macrophage dependent phagocytosis. Binding occurs through mannan polysaccharides present on surface proteins of the cell wall (Gauglitz *et al.,* 2012).

Reduced levels of pulmonary SP-D are observed during infection with *Coccidioides posadasii* in mice suggest that pulmonary SP-D may be involved in uptake and phagocytosis by antigen presenting cells and downstream immune regulation (Awasthi *et al.,* 2004). Coccidial antigens have been suggested to be the target of SP-D binding, but the exact mechanism has yet to be elucidated (Nayak *et al.,* 2012).

Saccharomyces cerevisiae aggregation was seen in the presence of SP-D. *S. cerevisiae*, like *C. neoformans* and *Histoplasma capsulatum* contains mannoproteins, but removal of these did not reduce binding. The presence of chitin and β-glucans present on the cell wall were assessed;

chitin and β1-3-glucan had no agglutinating effect, suggesting that β1-6-glucan is the primary site of SP-D binding (Allen *et al.,* 2001b).

1.4.5 Schistosome recognition

de Wetering and coworkers identified the fucose-containing ligands present on the surface of *S. mansoni* which are known to be bound by SP-D, and determined that the terminal residue is required to be in a specific conformation to be recognised.

It has been shown that SP-D binds in a calcium-dependent manner to F-LDN and F-Gn with relatively high affinity; no binding was seen with LDN, LDN-F and FF-Gn; and only weak binding was seen with $Lewis_X$ antigens (see Fig 1-15). The fact that the terminal fucose of FF-Gn was not bound by SP-D suggests that binding must occur via interactions with the equatorial hydroxyl pair O2 and O3 of the carbohydrate ring. Binding of $Lewis_x$ antigens, even as low affinity ligands, was suggested as a possible mechanism by which SP-D binds neutrophils and monocytes, and it is conceivable that fucose-containing glycoconjugates serve as endogenous ligands on leukocytes. Once bound to schistosomes SP-D causes agglutination by cross-linking or by attachment to cells in the alveoli, preventing further migration (de Wetering *et al.,* 2004b).

1.5 Aims and objectives

The atomic structure of a physiologically relevant bacterial cell surface carbohydrate bound to SP-D is yet to be solved, and the aim of this thesis is to successfully characterise SP-D binding to lipopolysaccharide of Gram-negative bacteria - and provide structural insights into longer oligosaccharide binding. This will be achieved via crystallisation of a biologically active recombinant head and neck fragment of human SP-D. Following the growth of reproducible, diffraction quality crystals, the goal is to soak an inner core region of an appropriate LPS into the crystal during the cryoprotection phase. Following data collection and structure solution, the result will be an SP-D electron density map which defines the binding determinants of the introduced ligand. The potential of recombinant SP-D therapy to reduce inflammation in neonatal lung disease, cystic fibrosis and emphysema has long been suggested. Insights into the exact mechanism of ligand binding may lead to enhanced understanding of the physiological interactions of SP-D, and thus open novel possibilities for potential therapeutics in a clinical setting.

Chapter 2: Protein Crystallography

X-ray crystallography provides highly detailed molecular structures, and as of March 2016, over 102,000 protein structures deposited in the Protein Data Bank have been elucidated in this manner (rcsb.org). Crystallography now provides the foundation of modern structural biology; whole viruses, ribosomes and gene-regulating proteins are but three examples of how crystallography has vastly improved understanding of the structural basis of biomolecular function. The quest for new drug therapies has invited momentous commercial and public interest, and crystallography plays a significant role in protein and drug-target discovery.

The first x-ray diffraction patterns were obtained in 1912 by Walther Friedrich and Paul Knipping from crystals of simple materials such as diamond and rock salt, and confirmed the atomic construction of matter and the interactions and bonds of the matter itself. The understanding of atoms and quantum mechanics were in their infancy, and these experiments provided the much needed evidence and support to develop the atomic theory of matter. Interestingly, it was already known that proteins could form crystals, but it would take another three decades before the first diffraction pattern of protein crystals would be recorded - and another two decades before the first macromolecular structure, myoglobin, would be determined.

It is often easy to take the tool of modern biomolecular crystallography for granted; how it is possible to tackle the task of elucidating structures the size of a single protein. We are spoiled by recombinant production techniques, robotic automation and powerful synchrotron radiation sources - the first protein crystallographers were pioneers and their efforts to obtain structures were no less than gallant. With good reason, the field of crystallography and structural biology is highly populated with Nobel Laureates.

Obtaining a molecular model from a protein crystal is deceptively simple: a single crystal is placed into a finely focussed x-ray beam, and diffraction images are recorded. Electron density that represents the atomic structure of the protein is reconstructed from the data, and an atomic

model is built into it. But both practical and fundamental challenges arise. For instance, well-diffracting protein crystals need to be produced. Proteins by their nature are large and flexible, and seldom readily self-assemble into regular, repeating arrangements typical of crystals. Producing pure protein and growing the crystals is often problematic, and biological material itself is susceptible to radiation damage during the diffraction experiment. In principle, crystallography poses challenges: it is important to recognise that it is not an imaging technique. X-ray diffraction patterns do not provide a direct image of the molecular structure and electron density must be reconstructed by Fourier Transform techniques, which require both amplitudes and relative phase angles of each reflection, the latter not being directly accessible from the data. Refinement of the structure is also theoretically demanding, and requires prior stereochemical knowledge to be successful.

One can readily appreciate that the crystallographic determination of protein structures is no mean feat. So why do it? Because it provides very detailed molecular information; atoms in a protein as well as atomic details of ligands, inhibitors, ions, and other molecules that are incorporated into the crystal.

2.1 Crystallisation

Determining the structure of a protein by x-ray crystallography requires growing high-quality crystals of a purified protein sample. Under certain circumstances proteins separate from solution and self-assemble into a periodic, repeating crystal lattice. A solution of sufficiently high concentration (usually 2-50 mg/mL, Dessau & Modris, 2011) becomes supersaturated with addition of precipitants that slowly reduce protein solubility. Once the solubility limit of the protein is reached and nucleation sites form (from which phase separation proceeds, either due to homogeneous formation at supersaturation or heterogeneous introduction of nuclei through seeding or foreign matter) crystals may start to grow.

Conditions that yield well-diffracting crystals are not easily predicted, and trials are often necessary - varying the reagent concentration, pH and temperature is required to optimise growth. The most popular manual technique for growing crystals is the *vapour diffusion* method, illustrated below via the sitting drop system. Chosen reagents are placed in a well (referred to as the *reservoir solution*), and microliter amounts of protein is added to an equal volume of the reagent mixture which is placed in a Micro-Bridge - a small inverted "U" made from clear plastic with a concave depression on top. A glass slide seals the well (see Fig 2-1).

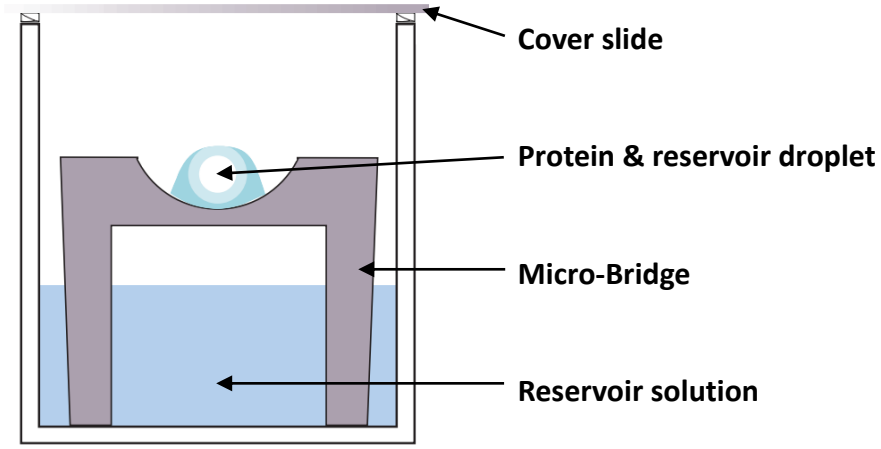


Figure 2-1 Diagram of the sitting drop method of protein crystallisation. The droplet is composed of an equal volume and protein sample and reservoir solution that, due to the sealed environment, undergoes vapour diffusion with a net movement of water from the drop - until the solubility limit of the protein is reached. Once a supersaturated state is achieved crystal nucleation may occur. Edited from Dessau & Modris, 2011.

The drop on the bridge contains a lower reagent concentration than the reservoir solution, and over time the net result is that water evaporates from the drop in to the reservoir; as partial pressure is higher in the droplet, water evaporates until equilibrium exists between the drop and the reservoir. As a result, concentrations of both protein and precipitant in the drop increase slowly until the protein's solubility limit is reached, and equilibrium is achieved when the reagent concentration of the drop equals that of the reservoir. The drop solution becomes supersaturated, nucleation and phase separation occur, and crystals may form.

2.1.1 Protein preparation

The development of techniques for producing and purifying protein has been essential for the advances made in structural biology. The ability of molecular biology to provide large quantities of proteins that are otherwise difficult to obtain ensures pure samples of target proteins are readily available. The objective of recombinant protein expression is typically to produce a sample that will illustrate a certain biological function, such as enzyme catalysis or protein-ligand interaction - and as such it is often not necessary to express the full-length protein. A bewildering array of methods may be employed in producing a protein recombinantly, including various choices of expression systems, host system, expression vector and affinity tags that can be used to simplify protein purification. The cells must then be harvested and the protein extracted and purified.

2.2 Crystal nucleation and growth

A practical way to represent the change of protein solubility with precipitant concentration is with a solubility diagram (see Fig 2-2). This visualises the general observation that the higher the precipitant concentration of a solution, the lower the maximal protein concentration and *vice versa*. The solubility lines in the diagram separate the regions of individual phases of the protein solution, and describe the entropic contributions that drive crystallisation. The energy barrier from the *undersaturated zone* where protein is in solution to *supersaturation* must be overcome in order to achieve nucleation. Once the activation barrier is overcome and nucleation sites are created, excess protein comes out of solution, and a protein-rich phase in equilibrium with the saturated solution occurs in the form of a crystal.

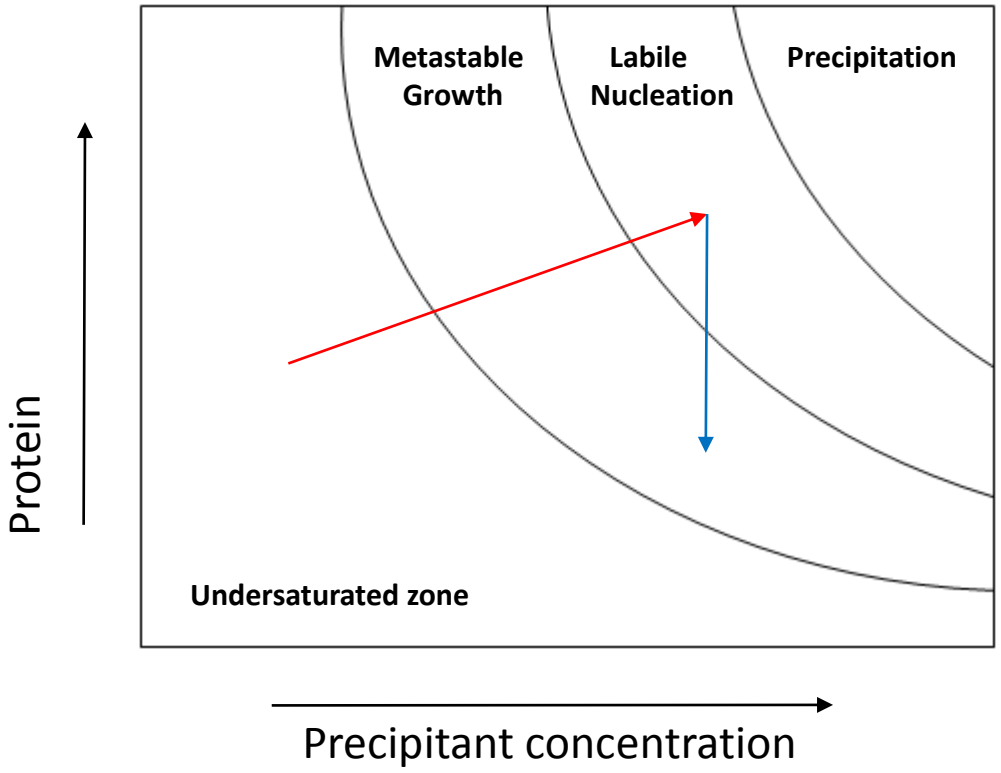


Figure 2-2 Protein crystallisation phase diagram of crystal growth via vapour diffusion, showing the solubility of protein in a solution as a function of the concentration of precipitant. The red arrow represents phase transition from protein in solution to nucleation, having crossed the nucleation energy barrier. The blue arrow represents growth of the crystal nuclei and subsequent drop in free protein concentration. Adapted from Asherie, 2004.

2.2.1 Nucleation

A protein is stable when net attractive interactions between the solvent and protein occur. Since the 19th century, it has been known that crystals form in solution by accretion, that is, the gradual accumulation of additional layers of matter - growth units freely moving in a solution attach to the crystal surface under appropriate conditions. Energetic forces contribute toward formation of stable crystal nuclei: self-assembly of proteins requires a net drop in free entropy to occur due to the significant loss in conformational freedom upon formation of crystal contacts, $\Delta S_{protein}$, around -25 to -75 kilocalorie per mole (kcal/mol).

This large loss of entropy results in a positive destabilising contribution to the crystallisation energy, ΔG_c, and so must be compensated by some other effect to render it negative, required for driving equilibrium from protein in solution to a crystalline solid. Enthalpic contributions, ΔH_c, have been shown to be weakly negative or practically insignificant for crystal formation (Derewenda & Vekilov, 2006) and as such the remaining stabilising contribution must come from elsewhere.

Disorganisation of the well-ordered water molecules that otherwise form clathrate cages round hydrophobic residues provides a strong entropic gain of the solvent, $\Delta S_{solvent}$ - the release of water molecules across polar and hydrophobic surface residues during the formation of crystals. Release of just 5-30 water molecules from the surface of protein during crystal formation is equal to the entropic gain in the range of 25-150 kcal/mol, and provides the remaining stabilising contribution required to render the total free energy of crystallisation negative.

The equation for the free energy of crystallisation is as follows, and describes a loss of entropy for the protein, and a gain in entropy for the solvent:

$$\Delta G_c = \Delta H_c - T(\Delta S_{protein} + \Delta S_{solvent})$$

It emphasises the fine balance between enthalpic gains and entropic changes during crystallisation, and how minute changes can upset the delicate equilibrium between a crystal and its surrounding solvent. Protein-solvent interactions compete with intramolecular interactions holding the protein crystal together, and one aims to shift the equilibrium toward increasing the latter gently, without denaturing the protein and driving it out of solution in the form of precipitation.

Reducing protein solubility allows us to move into the metastable region of the phase diagram (see Fig 2-2), but the occurrence of phase separation and self-assembly of molecules in to a crystal often requires some perturbation or activation event to overcome the kinetic barriers. In

the metastable, supersaturated region, nucleation must ensue and phase transition in to a new protein-rich phase occurs. Given that molecules in a supersaturated solution are rather concentrated, and collide frequently some of these impacts result in the formation of favourable contacts, and the consequential binding energy may overcome entropic loss of decreased order. This is referred to as *homogeneous nucleation*, and likelihood of stable nuclei formation in this manner increases with supersaturation. *Heterogeneous nucleation* on the other hand, is the artificial induction of nucleation and the basis of 'seeding' techniques whereby external crystallisation nuclei are introduced to a supersaturated solution to induce nucleation.

If supersaturation is too high, rapid growth of nuclei form a shower of tiny crystals to appear, and this occurs when the protein solution is maintained in the labile zone for too long. To obtain large, useable crystals (usually a few hundred µm down to roughly 10µm), a solution must approach the labile nucleation zone slowly.

2.2.2 Growth

When the solution is metastable, and nucleation has occurred successfully, it separates in to two forms: a protein-rich phase in the form of crystal clusters, and a saturated solution. This causes a shift into the *growth phase* where additional growth units accumulate in to the characteristic arrangement of a crystalline lattice. The actual growth of protein crystals can be observed through atomic force microscopy, whereby a microscopic needle of silicon nitride scans over a forming crystal surface, and deflections are measured by a laser while leaving the growth process undisturbed (McPherson *et al.,* 2003).

During nucleation, the effect of newly generated surface energy is minimised when new growth units attach themselves to ledges or steps in the crystal - which generally cause spiral-like growth patterns in crystals, see figure 2-3.

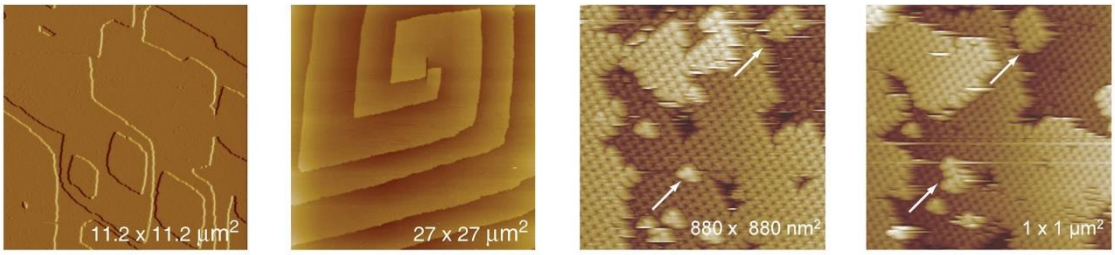


Figure 2-3 images of crystal growth using atomic force microscopy. The left panels show step- and spiral-growth patterns of crystal growth. The right panels show 2-dimension nuclei formation (indicated by arrows) and generation of a new nucleus from which a new growth-step develops. Images taken from Rupp, 2010.

These images also allow investigation into solution contamination and its effect on crystal growth. Because only the macromolecule of choice is stacked in to crystal form, contaminant concentration increases, and eventually inhibits further aggregation. New nucleation 'islands' may form on the crystal surface which may not perfectly align with the original crystal - a new, slightly misaligned domain forms (see Fig 2-4). If strongly misaligned nuclei inter-grow as one crystal, *macroscopic twinning* can be observed using a simple microscope. In this case, the crystal clusters can often be dissected into individual fragments (as they are single crystals in their own right), and can be used for diffraction. *Merohedral twinning* however, is not visually detectable. Domains alternate in orientation and generate diffraction patterns virtually indistinguishable from normal, abnormal only in their intensity distribution. A real crystal is rarely a single crystal, but rather a mosaic of near perfectly aligned domains. Misalignments of no more than a few tenths of a degree are useful for diffraction.

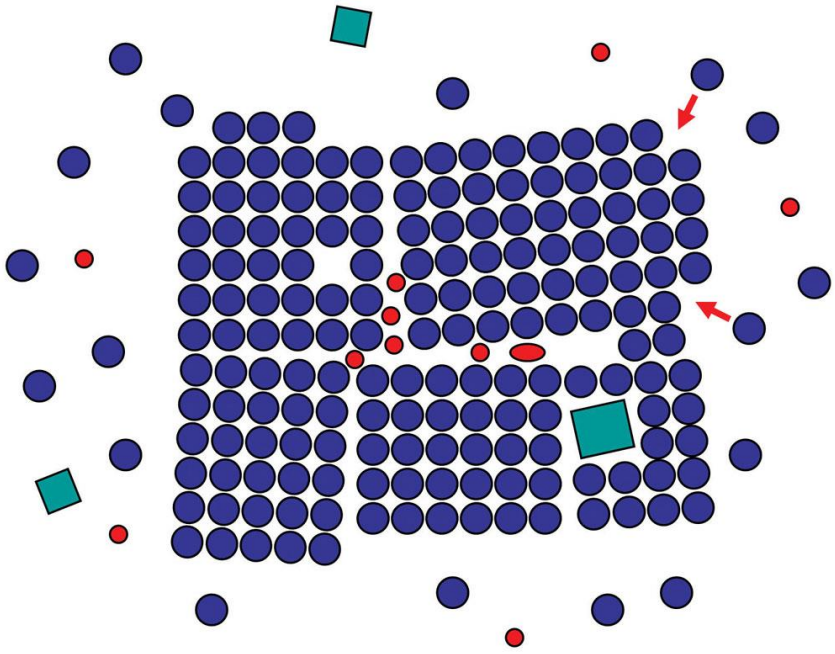


Figure 2-4 Schematic drawing of a highly mosaic crystal growing in solution. Protein molecule (blue spheres) are interrupted by small impurities (red) and debris (green). New molecules preferentially attach to ledges or steps, shown with red arrows. Contamination occurring within the growing crystal causes defects and misalignment, in this case of around 6°. Image taken from Rupp, 2010.

2.2.3 Creating a supersaturated state

A strategy employed to induce crystallisation is to gradually guide the protein toward a state of reduced solubility by modifying the properties of the solvent; this can be accomplished by either increasing the concentration of precipitating agents or by altering some physical property of the protein such as disturbing the protein's electrostatic requirements or minimising the dielectric shielding between potential bonding partners. Common approaches to reducing a protein's solubility include:

- Dialysis - as a means of changing the properties of the protein-containing solution, whereby the protein gradually acquires the desired properties of the exterior fluid: pH, ionic strength etc

- Vapour diffusion - the most common technique, as previously described, whereby evaporation of water from a protein droplet causes increased protein concentration
- pH adjustment and temperature adjustment - to perturb the protein's electrostatic requirements and alter the surface features or conformation to reach a solubility minimum
- Addition of precipitating agents - to form transient, nonspecific aggregates with the protein
- Addition of a ligand that reduces solubility - which may cause an increased degree of conformity amongst protein molecules (Hassell *et al.,* 2006)
- Addition of a polymer - such as poly(ethylene glycol) to reduce volume in the solution. Polymers lack consistent conformation and occupy large amounts of space due to large hydrodynamic radii. They compete with other macromolecules for available space
- Addition of a cross-linking agent - to bridge and stabilise intermolecular contacts between protein molecules (Yeh *et al.,* 2005)
- Removal of a solubilising agent
- Surface Entropy Reduction – replacement of high-entropy surface residues that may hinder formation of intermolecular contacts (Derewenda & Vekilov, 2006).

Any combination of these may be employed to refine the parameters of the system to ensure the slow phase transition from protein in solution to an aggregated cluster of molecules in the form of a crystal lattice. If supersaturation has proceeded too extensively or rapidly, the protein forms an amorphous precipitate; too slowly, and nucleation doesn't occur at all. It is almost always necessary to screen conditions to optimise crystal growth as each protein is unique, and few means are available to predict the conditions that will result in successful crystal growth.

2.2.4 Factors affecting crystallisation

The following table represents some, but probably not all, the physical, chemical and biological variables that influence crystallisation of proteins.

Table 2-1 Physical, chemical and biological variables that influence crystallisation of proteins

Physical	Chemical	Biological
Temperature	pH	Purity of sample
Surface	Precipitant type	Ligands
Method	Precipitant concentration	Aggregation state of molecule
Gravity	Ionic strength	Posttranslational modification
Pressure	Specific ions	Source of protein
Time	Degree of supersaturation	Proteolysis/hydrolysis
Vibration/sound	Redox environment	Chemical modification
Electrostatic field	Protein concentration	Genetic modification
Dieletric property of medium	Metal ions	Inherent symmetry of molecule
Viscosity	Cross-linking	Stability of molecule
Rate of equilibrium	Detergents or surfactants	Isoelectric point
Nucleation method	Impurities	Sample history

2.3 Data collection and measurements

The goal of data collection is to determine the indices and record the intensities of as many reflections as possible, as quickly as possible. When collecting data, many factors must be considered; sidechain conformation becomes clear at around 3 Å resolution but given the length of a carbon-carbon bond, individual atoms are resolved at around 1.5 Å. It is therefore desirable to collect data at the highest resolution achievable. According to Bragg's law, shorter interplanar spacings are resolved with higher scattering angles which are also associated with a fall-off of the form factor, *f*. This can be partially compensated for by increasing the intensity of the incident radiation.

For a successful diffraction experiment, a reliable source of x-rays is necessary. Three general types of x-ray sources are available: laboratory (or in-house) generators, which rely on emission of *characteristic radiation* from materials when electrons from outer shells fall back in to core level holes generated from electron bombardment; synchrotron sources, which generate x-rays by accelerating electrons at high energy; and Compton sources, which utilise high repetition lasers to excite oscillation in electron beams to generate x-rays. Unlike in-house generators in which anode elements emit radiation at specific wavelengths (chromium at 2.29 Å, copper at 1.54 Å or molybdenum at 0.71 Å) synchrotron radiation sources are 'tuneable' to a required wavelength - and they allow collection of data from very small crystals which would otherwise show little detectable diffraction.

Data for this thesis was collected at Diamond Light Source, the UK's synchrotron. The machine itself is an assembly of components that produce, confine and maintain high-energy electron beams in a closed orbit to produce synchrotron light. The Diamond machine is composed of the following main features:

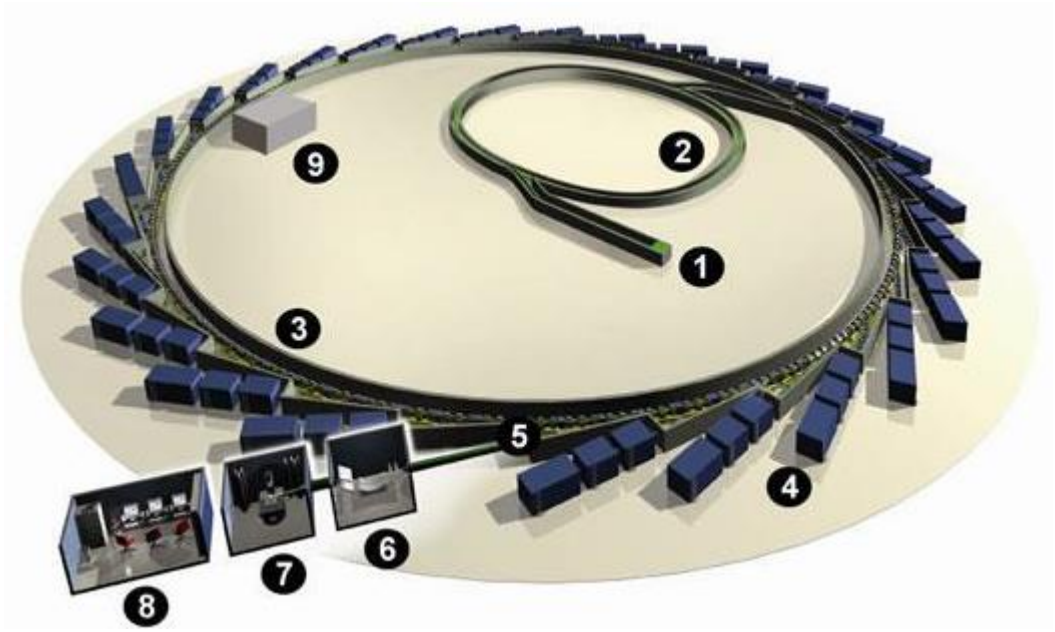


Figure 2-5 Birdseye view of the Diamond Light Source Synchrotron showing the nine components of the machine. In numerical order the components are: the injection system, booster synchrotron, storage ring, the beamline, front ends, optics hutch, experimental hutch, control cabin and radiofrequency cavity. Image adapted from diamond.ac.uk

1. An injection system. A high voltage cathode is heated under vacuum to give electrons in the material sufficient thermal energy to 'evaporate' and escape. They are accelerated by earthed anodes to produce a stream of electrons with an energy of 90keV. A linear accelerator accelerates the electrons to extreme energy of 100MeV using radiofrequency (RF) cavities
2. A booster synchrotron. Electrons follow a circular trajectory, along which dipole bending magnets curve the electrons round. RF voltage sources further accelerate the beam. The magnetic field produced by the bending magnets increases up to a maximum of 0.8 Tesla. The electron energy is raised to a final 3GeV
3. A storage ring consisting of twenty-four straight sections, angled to form a closed loop, produce a closed ring of over 560 metres. The entire ring is maintained under vacuum to minimise electron scattering by particles and atoms in air

4. Beamlines. The research station, which in itself consists of the four following areas (hutches)
    - Front ends. Where synchrotron light is channelled into a beamline. It monitors the light, and protects the storage ring in case of leaks, as well as removes heat to allow safe access to the optics hutch
    - Optics hutch. The first room that synchrotron light enters, where the beam is filtered and focussed
    - Experimental hutch. Where the experimental equipment is housed. The crystal is placed in the x-ray beam on a rotating, robotic arm which is surrounded by detectors
    - Control cabin. The final hutch of the beamline. Where the experiment is monitored - the alignment and position of the crystal can be controlled, as well as specifics of data collection
5. A radiofrequency cavity. As electrons circulate in the storage ring they pass through bending magnets and insertion devices which radiate energy in the form of synchrotron light. The cavity provides an additional energy boost to compensate for energy losses, and maintain a fixed orbit round the storage ring.

Once x-rays have been diffracted by a crystal, they need to be efficiently detected. Crystallography experiments utilise area detectors which cover large solid angles of diffracted radiation. Examples include x-ray sensitive film and imaging plates; multiwire array detectors in which x-ray photons are absorbed by a beryllium cathode and ionise a gas which generate an electronic pulse that can be recorded; charged-coupled devices (CCDs) whereby a semiconductor array transforms radiation in to an electric signal; phosphor-based, analogue Complementary Metal-Oxide Semiconductor (CMOS) detectors; and more recently, hybrid pixel array detectors, such as PILATUS which employ an array of silicon sensor pixels that directly detect x-ray photons of a desired energy.

In order to collect a set of diffraction data, the entire asymmetric unit of reciprocal space must be sampled, which requires rotation of the crystal round at least one of its axes, denoted φ. The instrument that facilitates this is the *goniostat* or *goniometer*, which also allows translational movement in *x*, *y* or *z* to centre the crystal in the beam. All data sets presented in this thesis were collected on synchrotron beamlines at Diamond Light Source Ltd, Diamond House, Harwell Science and Innovation Campus, Fermi Avenue, Didcot, Oxfordshire OX11 0QX, United Kingdom.

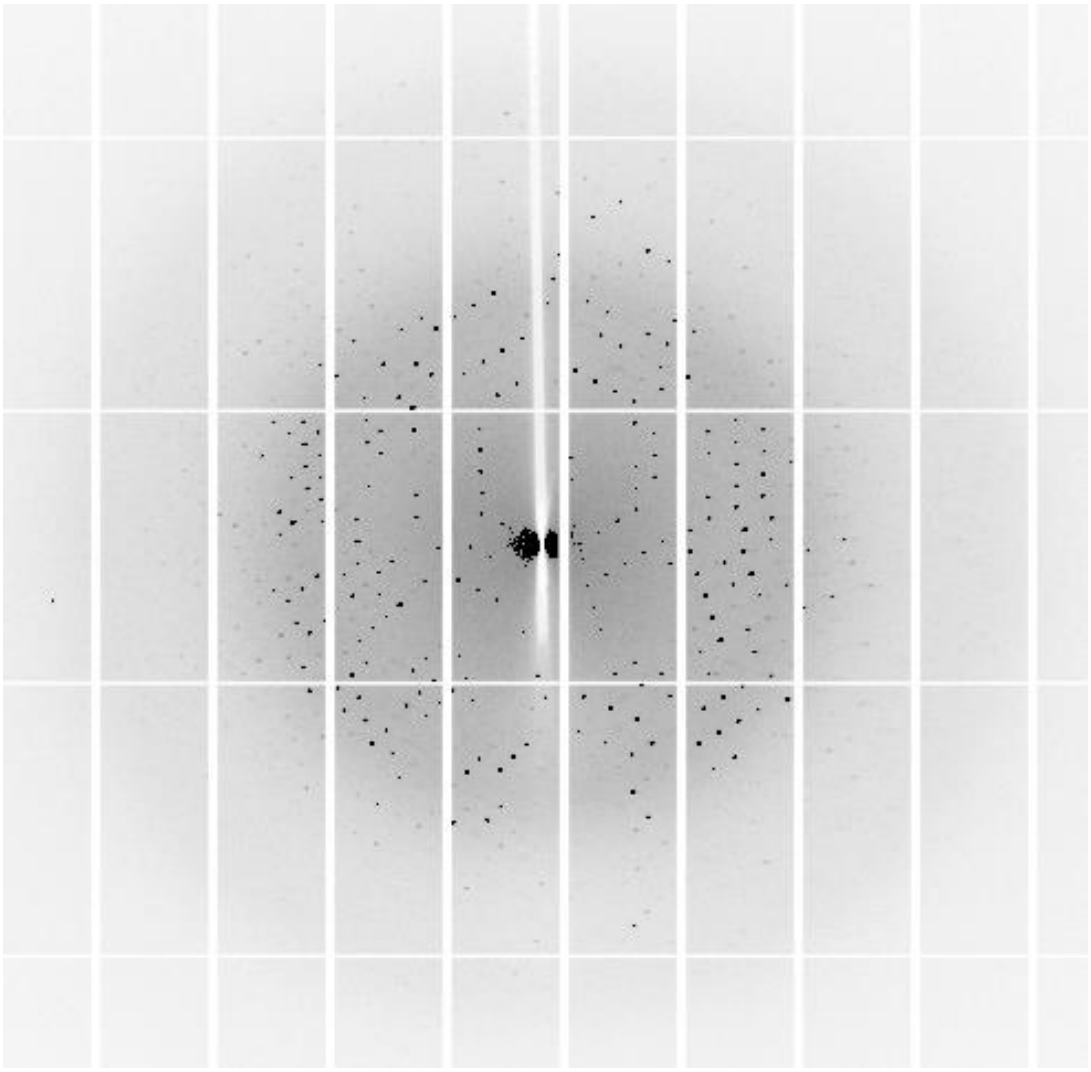


Figure 2-6 Diffraction pattern from a crystal of recombinant Surfactant Protein-D. Reflections form a regular pattern but intensity (spot darkness) varies, and fade toward the outer regions of the image - corresponding to the data set's resolution. Image collected on MX beamline I03 at Diamond Light Source.

2.3.1 Radiation damage and cryoprotection

A cause of urgency during data collection is that crystals deteriorate in an x-ray beam due to the heat and reactive free radicals that are generated. A vital tool in modern crystallography is the availability of cryocooling equipment during data collection to maintain crystal integrity. It has been shown that lowering the temperature of protein crystals improves the lifetime of the crystal during x-ray irradiation - improvements of at least 1000-fold can be obtained by cooling crystals to liquid nitrogen temperatures.

When crystals have grown to an appropriate size, they need to be harvested from the crystallisation droplet and mounted on the diffractometer. They do however, need to remain in the mother liquor to prevent drying out and eventually disintegrating. There are two methods of mounting crystals: *capillary mounting* where the crystal (and solution) are sucked up in to a thin-walled, quartz capillary tube, the solution is wicked out and the tube itself is sealed with wax then centred on a goniostat; and *cryomounting* whereby a single crystal is scooped out of the droplet with a fine nylon loop, then directly placed into liquid nitrogen - a process referred to as *flash-freezing*.

Figure 2-7 Photograph of a single crystal in a harvesting loop following cryocooling. The crosshairs indicate the crystal location and are typically used to position the crystal in the beamline during data collection. Image adapted from Watts *et al.,* 2010.

Freezing crystals while they are still in their aqueous mother liquor causes formation of ice crystals which can disturb the protein crystal lattice, lessen diffraction quality and cause undesirable rings, arising from ice crystals scattering the x-rays, on diffraction images. This process can be overcome in three ways: by rapidly freezing the crystal, on a timescale too fast for ice formation to occur; cooling at high-pressure to circumvent formation of typical, hexagonal ice forms; and by modifying the physiochemical properties of the solvent prior to flash-freezing by addition of a *cryoprotectant* to replace the majority of the water in the solution. The latter is the most common method employed by crystallographers, and typical choices of cryoprotective

agents include glycerol, ethylene glycol, 2-Methyl-2,4-pentanediol (MPD), sucrose and low molecular weight PEGs, which should be in a buffer compatible with the mother liquor. Unfortunately, it is not possible to predict which cryoprotectant will be suitable for a specific protein, and screening is generally necessary (Řezáèová, 2002).

2.3.2 Indexing and merging

Once images are collected computational programs such as iMosflm within the CCP4 suite (Battye *et al*., 2011) are used to find the diffraction spots on the images and index them. Indexing is the assignment of the reflections in terms of the reciprocal lattice parameters a*, b* and c*, which correspond to unit cell dimensions in real space. Depending on the quality of the data, it may be possible to index the diffraction pattern from a single frame. Multiple possibilities exist when defining lattice vectors this way, for instance any lattice can be indexed with arbitrary triclinic vectors $a \neq b \neq c$, $\alpha \neq \beta \neq \gamma$. Indexing programs match all 44 possible combinations of crystal systems plus Bravais centerings and rank them by a penalty function. After initial indexing, a data collection strategy is devised and gives suggestions of rotation ranges suitable for maximal coverage of the asymmetric unit in reciprocal space. It is seldom a disadvantage to collect redundant data, and given the high-throughput capability of synchrotrons and inherent symmetry of crystals, general strategies suffice to obtain complete data. It is possible at this point to estimate the *mosaicity* of a data set, a measure of the spread of crystal plane orientations. As seen in Fig 2-6 a real crystal is rarely a single crystal, but rather a mosaic of near perfectly aligned domains. Misalignments of no more than a few tenths of a degree are useful for diffraction and as long as the mosaicity value is lower than the rotation angle, integration of the data is unimpeded. Partial reflections (that is, spots recorded partially on more than one successive image) can be summed appropriately, but in some cases high mosaicity or very long unit cell axes cause reflections to overlap on the images which renders the data more difficult to interpret.

Once diffraction images have been indexed and a strategy deduced, the cell constants and experimental parameters are refined using 5-10 images. Although the unit cell is determined as part of the indexing process, it is possible to obtain more accurate measurements using a procedure that requires *integration* of the data, ideally from two or more separate data segments at different φ values. Integration combines partial reflections from multiple frames in to a single reflection, and multiple cycles are often required for convergence. If the shift in unit cell parameters is greater than 2.5 standard deviations, another cycle of refinement and integration is performed.

The end result is integrated raw data - essentially long lists of indices and intensities of each reflection that must be reduced and scaled.

Data reduction begins with the merging and scaling of multiple measurements from identical reflections, dependent on the initial indexing, to give a reduced data set that contains all reflections. Intensity of the data becomes weaker at high resolution due to fall off of scattering, and merging statistics are provided at this point for the entire data set, as well as binned resolution shells.

An increased error in the measurement of weak reflections corresponds to a poor signal to noise ratio, expressed as average $<|I|/\sigma(I)>$, summed over all N reflections in a resolution shell, where I(hkl) is the average of symmetry related observation of a unique reflection:

$$< |I|/\sigma(\mathrm{I}) > = \frac{1}{N}\sum_{h}^{N}\frac{|I_{\mathrm{h}}|}{\sigma(I_{\mathrm{hkl}})}$$

The most commonly used quality indicator when merging reflections within a resolution range is the *linear merging R-value*, or R_{merge}:

$$Rmerge = \frac{\sum_h \sum_{i=1}^{N} I_{(h)}i - \bar{I}_{(h)}}{\sum_h \sum_{i=1}^{N} I_{(h)}i}$$

where the inner summation extends over all N redundant observations for given reflection h, and $\bar{I}(h)$ is the averaged intensity. The outer summation extends over the desired resolution range. R_{merge} measures the differences between the same reflections measured at different points in the data collection. As a consequence of increasing relative error in the measurement of weak reflections, merging R-values increase rapidly with resolution. The lower the value, the more agreement. R_{sym} describes merging of symmetry-related reflections.

As a general rule of thumb, quality indicators of high data quality should include a signal to noise ratio, $\langle |I|/\sigma(I) \rangle$ of roughly 2 at the limit, a low R_{merge}, completeness that allows full sampling of the reciprocal unit cell, as well as a resolution that gives enough definition for accurate map fitting.

2.4 Macromolecular structure solution

Determining the structure of a protein from its diffraction pattern is a matter of converting the reciprocal space information in the pattern itself to that of real space inside the unit cell. We understand from Bragg's equation that the direction of reflected x-rays relates to the direction of the corresponding reciprocal lattice vector; with the unit cell and crystal orientation defining where each reflection *hkl* will be collected on the detector.

As previously mentioned, the intensities of all *unique* reflections are collected in a diffraction experiment to reconstruct the electron density of a molecule, that is a reduced set of recorded reflections that belong to one and the same asymmetric unit of reciprocal space. The next step is

to define the relationship between unit cell contents and the intensity of each and every one of the diffracted x-ray maxima.

2.4.1 Calculating electron density maps

To produce an image of a molecule from the crystallographic data, the electron density within the unit cells must be calculated from a list of indexed intensities. Fourier transforms are a mathematical relationship between the contents of the crystal unit cell and its corresponding diffraction pattern. As each reflection can be described by a structure factor equation, containing information for each atom in the unit cell, the electron density can be described as a Fourier series in which each term is a structure factor.

F_{hkl} can be written as a sum of contributions from each volume element of electron density in the unit cell. The electron density of a volume element centred at *x*, can be taken as the average value of ρ(*x*) in that region. In effect, the volume elements can be made infinitely small, and the structure factor can be thus expressed in terms of electron density and the reciprocal lattice as:

$$\mathrm{F}(\mathbf{h}) = \int_{cell} \rho(x)\exp(2\pi i \mathbf{h}\boldsymbol{x})\boldsymbol{dV}$$

where ∫ is the integral over all values of x, y and z in the unit cell. Each volume element contributes to F_{hkl}, with a phase determined by its coordinates.

2.4.2 Electron density maps

The final result of data collection is a molecular model of the unit cell contents, using measurements from the native data sets as well as calculated phases. At this point, the phases are usually rough estimates and so only a crude model can be discerned. Once a model is built in to the density map it is used to calculate new, more precise structure factors and add more terms to a Fourier series description of ρ(*x,y,z*). This produces a clearer map and allows more accurate location of atoms, which in turn produces more accurate phases and calculated structure factors.

2.4.3 Phase improvement

Phase improvement requires modification of the model, through methods such as solvent flattening, histogram matching or introduction of a partial atomic model. It is also possible to adjust the atomic coordinates to improve agreement between amplitudes calculated from the current model and those observed in the "native" data set. The outcome of phase improvement is to add information to the map, while simultaneously reducing noise.

Solvent flattening exploits the fact that disordered solvent regions in the map are essentially flat at about 0.33 $e^-/Å^3$, compared to protein which has an average electron density of roughly 0.44 $e^-/Å^3$. A necessity of electron density maps is that the protein map must be of sufficient quality to allow solvent regions to be distinguished from the molecule, and a *solvent mask* to be generated. Density modification programs such as DM (Cowtan, 1994) set a density cut-off value: density above the cut-off such as protein is kept, while density below the cut-off is removed. New structure factors are calculated using both the protein map and newly flattened solvent to give an improved map.

In addition to the aforementioned phase improvement methods, the presence of more than one copy of an asymmetric unit can advance phase improvement if it is assumed that each crystallographically independent subunit is near-identical. This is referred to as *noncrystallographic symmetry averaging* and is not always advisable at high resolution, as there may be slight variations between independent molecules arising from crystal packing, or in symmetry which is often vital for the protein's function.

At some point the electron density map becomes clear enough that a protein chain can be traced through it and defined structural motifs such as alpha-helices can be recognised. Graphical interfaces such as O (Jones *et al.,* 1991) and Coot (Emsley & Cowtan, 2004) allow the construction and manipulation of stick-models of known sequence in to the map, *map fitting*. From the resulting model, new structure factors are calculated using phases which contain information

from the new molecular model rather than an approximation of ρ(*x,y,z*). Additional structural components are then added to improve the map further; the new, more accurate structure factors improve phase and density become clearer.

2.4.4 F_O and F_C maps

With each successive map new molecular features are added as they are defined, and errors in the model such as side-chain conformations are corrected. As the model nears completion difference maps are used to distinguish any remaining disagreements between the model and the data. Two distinct parts of the Fourier synthesis are required for electron density reconstruction: structure factor amplitudes, and phase. Electron density reconstructed using calculated structure factor amplitudes as the coefficients in the electron density equation and calculated phases is referred to as the F_C map; while maps calculated from the observed structure factor amplitudes and calculated phases are referred to as F_O maps.

By selecting a combination of map coefficients the interpretability of the electron density map is greatly improved. For example, a *difference density map* of F_O-F_C indicates errors in the model by showing areas of positive density where the model has missing elements which are present in the protein structure as well as areas of negative density which indicate areas that should have no model. Regions where the map shows no contours represent agreement between the model and actual density.

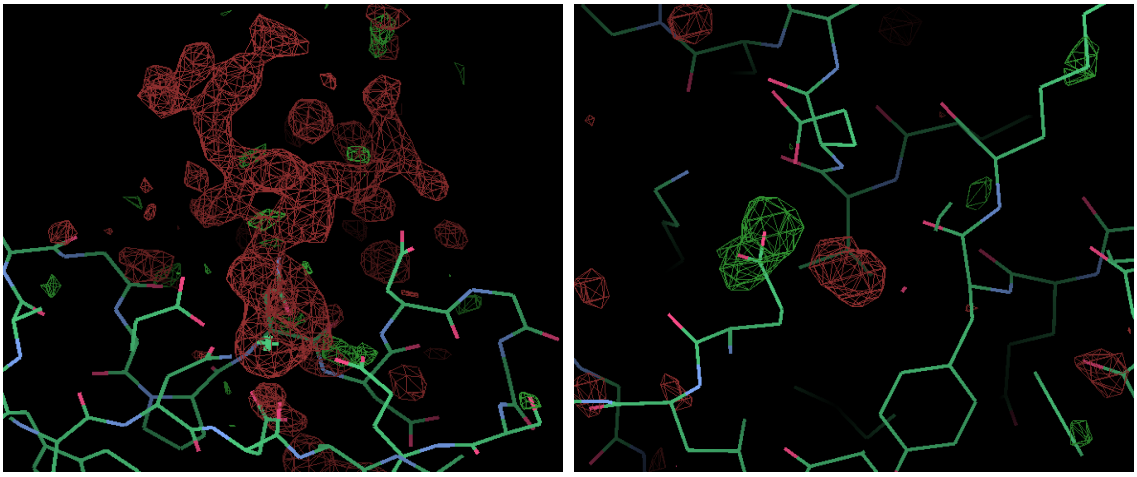


Figure 2-8 F_O-F_C map showing a large area of positive density, red, corresponding to a missing ligand (left), and an area of negative density, green, indicating a conformational irregularity of an amino acid sidechain.

In cases where a ligand has been soaked in to a macromolecule, areas of positive density may indicate the ligand's presence and position (see Fig 2-8, left panel). The absence of density in areas that are correct makes F_O-F_C maps suitable for detecting errors in the model or finding missing density, but not for initial model building. In this case $2F_O$-F_C maps are used. A $2F_O$-F_C map can be imagined as a summation of the difference map and the F_O map, and is useful not only for visualisation of the model (see Fig 2-9), but also for amplifying positive density during the early model building stage. Incomplete models that have missing atoms or other errors are reconstructed at half the density or less for the missing atoms. Using $2F_O$-F_C as a Fourier coefficient, the missing atoms are sigma-weighted up to normal density levels, allowing for easier model adjustment (Read, 1990).

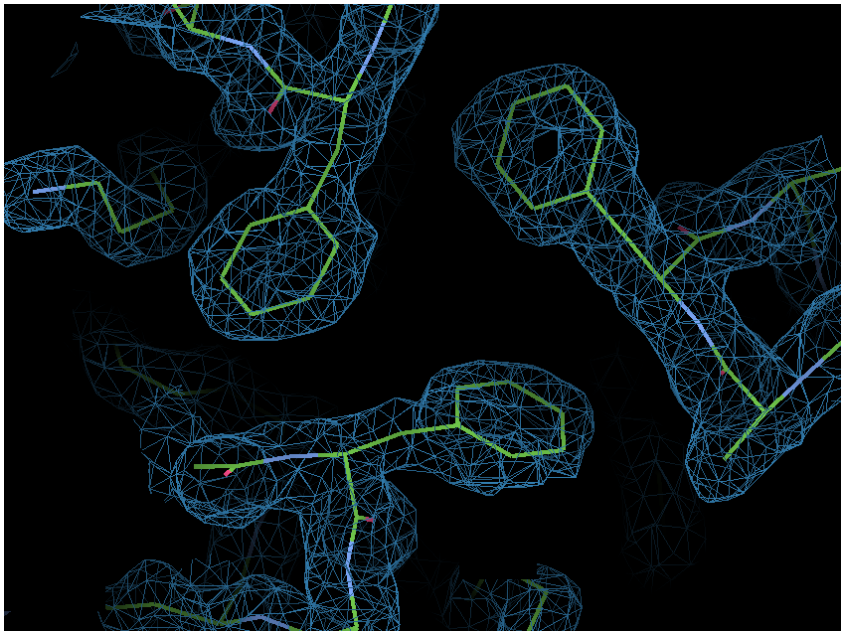


Figure 2-9 A $2F_O$-F_C map used for initial model building. It shows the current, best estimate of the electron density for the structure

2.4.5 The final model

Specific details of the structure are crucial for understanding the protein's mechanism, as ligand binding or functionally significant conformational changes may present as only small areas of density in the map. It is therefore necessary to construct the model carefully to provide as clear and precise a final model as possible.

2.4.5.1 Model building and refinement

Building a protein model in to an empty electron density map is generally unproblematic when the map has been reconstructed from accurate phases and high-resolution data, largely because side-chain orientations become distinct and the protein can be readily positioned. Manual model building starts with placement of Cα atoms into the density to form a Cα trace, then additional groups and sidechains are added to assemble a larger fragment. Orientations are adjusted to suit the unique electronic and spatial environment. Ligands are built in to areas of negative density in the same way. Map fitting becomes a matter of searching for and correcting errors in the model,

and the final structure can be assessed by comparison to universally valid stereochemistry, or by comparison of F_O and F_C - called *refinement*. Programs such as REFMAC (Murshudov *et al*., 1997) and SHELX (Sheldrick, 2010) calculate differences in values that should be equal and express this as a percentage called the *residual index* or *R-factor*:

$$R = \frac{\Sigma \, ||\mathrm{Fo}| - |\mathrm{Fc}||}{\Sigma \, |\mathrm{Fo}|}$$

The value of R ranges from 0 for perfect agreement of calculated and observed intensities, to around 0.6 for total disagreement. An R-factor greater than 0.5 is said to be random, and will not respond to improvement unless more data is available or a new model.

In addition to monitoring R-factors as an indicator of convergence, other structural parameters indicate whether a model is *chemically*, *stereochemically* and *conformationally* reasonable. Each atom in the three-dimensional structure has four distinct values: positional coordinates x, y and z and a temperature factor, *B*, which defines its atomic displacement from its mean position. For a chemically reasonable model the bond lengths and angles must be near the expected values for organic molecules, and a well-refined model exhibits a root-mean-square deviation of no more than 0.02 Å for bond lengths and 2° for angles; A stereochemically sound model has no accidentally inverted chiral centres (for example, D-amino acids); and a conformationally sensible model ensures that peptide bonds ψ (between N and Cα) and ϕ (between Cα and C) are within allowed ranges and torsion angles between bonds lie within a few degrees of stable conformations. Numerous rounds of refinement are performed, after each of which the above parameters should continually improve.

2.4.5.2 Validation

Limited resolution and imperfect phase information make building and refining a protein structure from diffraction data a somewhat subjective process, and errors in the final model are almost unavoidable. The crystallographer's task is to recognise and remove as many of these

faults as possible, prior to publication and deposition of a structure. With high-resolution data and high-accuracy phases, the resulting model is still a consequence of the data, and subjective decisions must be made - such as which refinement program to use, whether to include alternative conformations, whether to include explicit hydrogen atoms, which restraints to apply and so on. Although the purpose of model building and refinement is to detect and fix errors and obtain the best possible model, additional quality indicators are useful for final structure validation. The following list outlines various methods of model proofing.

- Ramachandran main chain torsion angle plots - representation of dihedral torsion angles of each residue in an energy contour plot, representing potential surface energy. Repulsive van der Waals interactions limit the possible torsion angles to certain regions. As main chain torsion angles are generally not restrained during refinement, torsion angle analysis provides valuable stereochemical validation (example Ramachandran plot shown below).

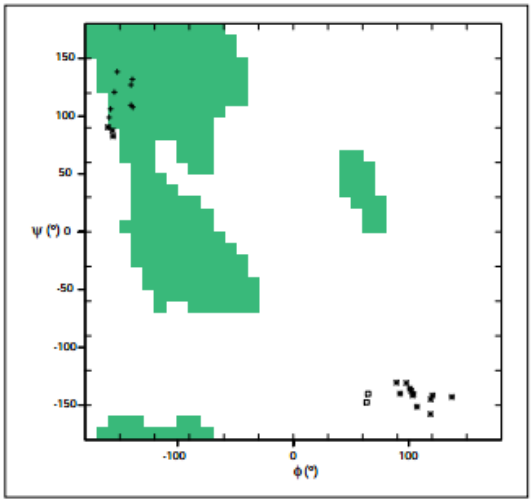


Figure 2-10 A Ramachandran plot. Glycine residues are shown as squares (bottom right of plot). Non-glycine residues are shown as crosses (top left of plot) if they fall inside core regions or as asterisks if they lie outside. Core regions are shown as the green areas (Kleywegt & Jones 1996).

- Kleywegt plot - examine differences between the torsions of NCS-related chains
- Geometry analysis - A restraints-based geometry analysis of the molecule. Checks for improbable bond lengths, angles, etc. against universally valid stereochemistry
- Chiral volumes - a dictionary is used to identify the chiral atoms of each of the model's residues. Checks for chiral centres with the wrong handedness.
- Rotamer analysis - checks for unusual protein side-chain conformations.
- Density fit analysis - identifies parts of the model which don't fit the density
- B-factor outliers - detects incorrect orientation of amino-carbonyl groups of glutamine and asparagine residues
- Unmodelled density/blobs - check for electron density not accounted for by existing atoms. For example, sulphates, ligands and misplaced sidechains
- Difference map peaks - one of the fastest ways to validate a model and its data. It highlights regions where the model and the data do not agree by more than 2 Å
- Check waters - Sometimes waters can be misplaced or imagined - taking the place of sidechains or ligands (Emsley *et al.,* 2010)
- MolProbity probe clashes - checks for hydrogen atoms with inappropriate environments (Chen *et al.,* 2009).

Once the molecule has been validated and thoroughly checked, one can begin characterising the protein and any potential ligand-binding. When the crystallographer chooses to do this, however, is a matter of preference and perseverance. The philosopher Paul Valéry once remarked that "a work of art is never finished, only abandoned" and he could well have said the same for structural models of macromolecules (Terwilliger & Brerendzen, 1996).

The following flow chart details the processing steps required to successfully determine a protein's structure from crystal data:

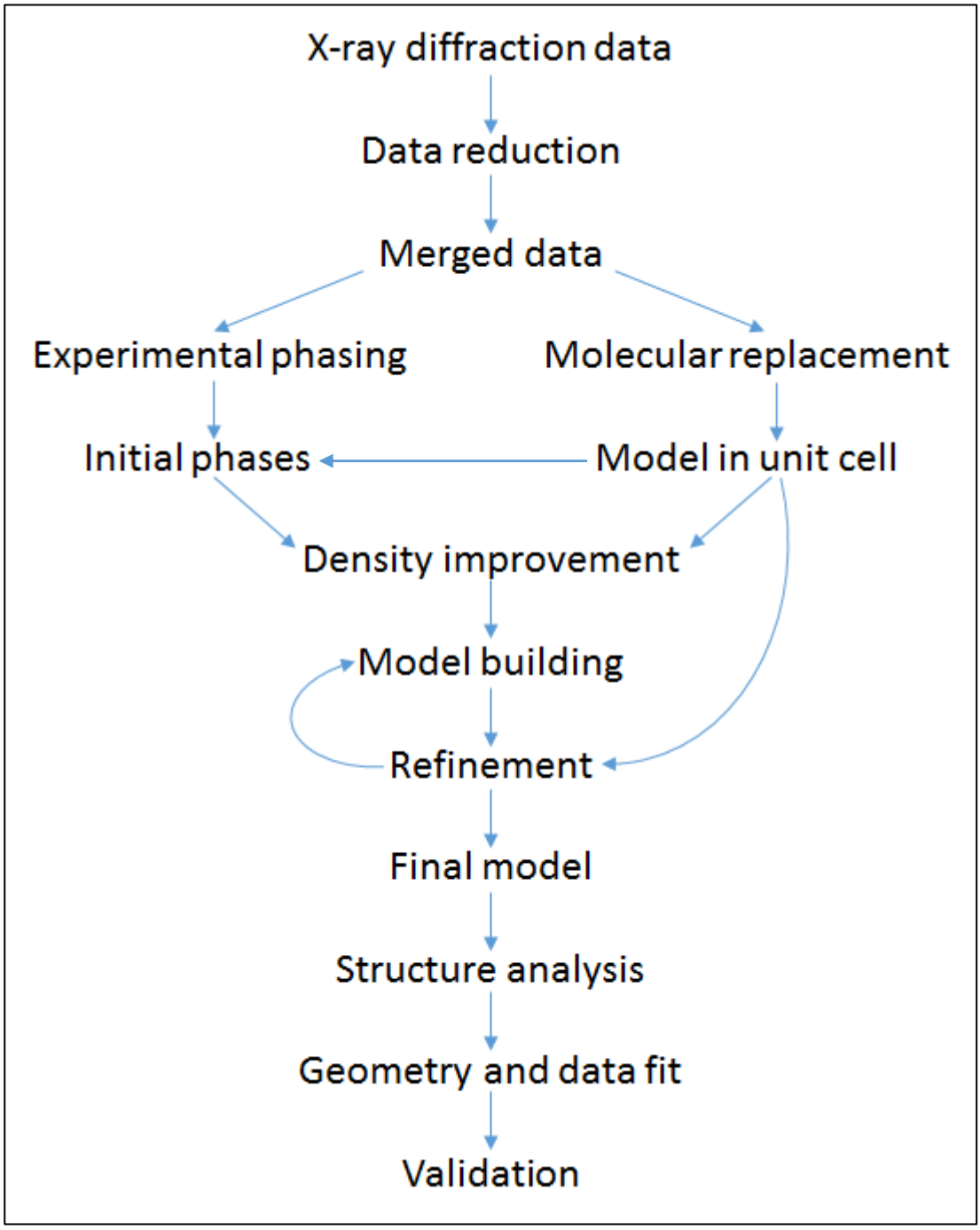


Figure 2-11 Flow chart showing the processing steps from data collection to model validation. Adapted from Potterton *et al.,* 2003.

Chapter 3: Experimental techniques

3.1 Protein

Three protein preparations of a trimeric recombinant head and neck fragment of human SP-D were generously donated by Jens Madsen and Howard Clark, University of Southampton, each having undergone different purification techniques:

1. Affinity purified, 0.5ml, 11.29 mg/ml (in TBS and 5mM EDTA)
2. Affinity purified and size exclusion chromatography, 0.45ml, 10.97 mg/ml (in TBS and 5mM EDTA)
3. Affinity purified, size exclusion chromatography and endotoxin treatment, 0.6ml, 8.76 mg/ml (in PBS)

3.1.1 Recombinant expression

A recombinant homotrimeric fragment of human SP-D was expressed using plasmid pUK-D1 containing cDNA for the neck and CRD (177 residues: Gly179 to Phe355) under bacteriophage T7 promoter in *E. coli* BL21(λDE3) pLysS. The expression cassette included a short stretch of eight non-terminal Gly-X-Y repeats corresponding to the collagen region (Gly179 to Asp203), with a substitution of serine for proline at position 2 (residue 180); followed by an α-helical coiled coil neck region (Val204 to Pro235); and a globular carbohydrate-binding head domain (CRD) (Asn236 to Phe355). The protein was expressed as inclusion bodies containing insoluble recombinant SP-D, which was refolded via a denaturation and renaturation procedure, of dialysis against decreasing concentrations of urea (6m to zero), which was then purified to varying extent, as described above. The protein purity was judged to be pure by SDS-PAGE, Western blotting and amino-terminal sequencing - and biological activity ability was examined by binding to simple carbohydrates (maltose, glucose, mannose, galactose among others), phospholipids and maltosyl-BSA. The recombinant protein behaved as a homotrimer of roughly 60kDa when examine by gel filtration chromatography, and under reducing conditions ran as a monomer of roughly 20kDa. No

higher-order multimers were seen under non-reducing conditions, confirming that oligomerisation was not a result of disulphide linkages between CRD domains.

3.1.2 Affinity chromatography

Affinity chromatography separates macromolecules on the basis of a reversible interaction between a protein and a specific ligand which is covalently bound to a chromatography matrix. The immobilised ligand retains its specific binding affinity for the target molecule, and subsequent protein-ligand interactions may be as a result of hydrophobic interactions, van der Waal's forces and hydrogen bonding. The target protein is then eluted from the medium by use of a competitive ligand, or non-specifically by changing the pH, ionic strength or polarity which reverses the interaction. When removed, the protein remains in active form. Affinity chromatography is also used to remove specific contaminants, such as serine proteases that lack binding affinity for the media, and allow many separations to be achieved in one step (Urh *et al.,* 2009).

3.1.3 Size-exclusion chromatography

Purification of protein by size-exclusion, also referred to as gel filtration, is a chromatographic technique in which macromolecules are separated by size. Separation is achieved by the differential exclusion of the sample solution by the pores of a matrix material. The medium consists of spherical, inert beads that form pores of a specific size distribution - and proteins separate depending on their ability to be included or excluded from them. Smaller molecules diffuse into the pores and their flow through the matrix is determined by their size - as they pass through they are eluted in order of decreasing molecular weight; while larger molecules do not enter the pores and are eluted in the column void volume. Unlike other forms of chromatography, size-exclusion chromatography does not cause molecules to bind to the column media, meaning that peak resolution is minimally affected by buffer composition (Mori & Barth, 1999).

3.1.4 Endotoxin treatment

In addition to isolation and purification of the macromolecules themselves, endotoxins liberated by Gram-negative bacteria cause frequent contamination of protein solutions derived from recombinant production. A general method of endotoxin removal is currently not available; the heat-stability and composition of LPS molecules (a hydrophobic lipid A moiety and hydrophilic complex carbohydrate region with negative charged phosphate groups) provide unique features with respect to possible interactions with other molecules. Techniques to remove endotoxins include:

- Adsorption - whereby solutions are passed through a column containing high-affinity adsorbents such as immobilised L-histidine, poly-L-lysine and peptide antibiotic polymyxin-B. Contaminant LPS binds to the matrix and purified protein is eluted.
- two-phase partitioning - whereby an aqueous surfactant solution spontaneously separates into two liquid phases, one of which contains a greater concentration of LPS micelles (aggregation of molecules formed by non-polar interactions of the alkyl chains of lipid A and surfactant tail groups) that can be removed
- Ultrafiltration - which removes large endotoxin aggregates, but this is not effective if LPS exist as monomers or interacts with protein molecules
- Chromatography - such as ion-exchange chromatography, a common method of endotoxin removal in protein preparations, enables the negatively charged phosphate groups on LPS molecules to interact with anion exchange resins on a column. Non-bound protein is eluted (Magalhães *et al.,* 2007)
- Kits are also widely available that contain an endotoxin removal resin of porous beads that bind selectively to endotoxins.

3.1.5 Dialysis

Dialysis is a common method of diffusion used to remove small, unwanted molecules from a protein solution and for buffer exchange of a protein sample. A membrane of suitable molecular weight cut-off facilitates movement of molecules in solution from areas of higher to lower concentration until an equilibrium is reached (Phillipis & Signs, 2005). The protein preparations used in this thesis underwent dialysis in a Medicell Visking nitrocellulose dialysis membrane with a 12,000-14,000Da molecular weight cut-off to remove EDTA from the sample, as well as for buffer exchange in to 10mM Tris, 10mM $CaCl_2$ and 140mM NaCl at pH 7.5 prior to crystallisation trial setup.

3.1.6 Concentrating and NanoDrop

Concentration of dilute protein samples was achieved through centrifugation with Merck Millipore and Amicon Ultra centrifugal filter units with a molecular weight cut-off of 10,000Da. Preparations underwent centrifugation until a chosen volume was achieved, and post-dialysis protein concentrations were tested using a Thermo Scientific NanoDrop 1000 Spectrophotometer. 1µl samples of a protein solution were placed on the device pedestal and an absorbance reading at 280nm was taken, which along with a user-input extinction coefficient of 0.494 moles per centimetre (based on a protein's atomic composition, and in particular the presence of aromatic amino acids such as tyrosine, tryptophan and phenylalanine which exhibit strong UV-light absorption), to give an accurate protein concentration. Dividing the absorbance of a sample solution by the molar extinction coefficient gives the molar concentration of the protein solution (Gill & von Hippel, 1989).

3.2 Ligand preparation

The hydrophobic lipid A moiety of LPS renders the molecule insoluble, and soaking a ligand into a crystal require solubility. Lipid A is also large in size, roughly 30 Å by 30 Å (Qiao *et al.,* 2014) and as

such often requires removal before ligands are able to enter crystal solvent channels. The bond between Kdo and the inner glucosamine of the lipid A is very acid labile, and is typically cleaved under mild conditions.

Lipopolysaccharide from *Haemophilus influenzae* Eagan 4A was generously donated by Derek Hood and Mary Deadman, Univerisity of Oxford, UK; Lipopolysaccharides from *E. coli* 0111:B4 (L3012), *E. coli* 026:B6 (L2762), *E. coli* J5 (L5014) and *E. coli* F583 (L6893) were purchased from Sigma-Aldrich; Lipopolysaccharide from *S. minnesota* R7 (ALX-581-018-L002) was purchased from Enzo Life Sciences.

3.2.1 Mild acid hydrolysis of lipid A

The method of removing the lipid A moiety from LPS was adapted from Phillips *et al.,* 1992, whereby intact lipopolysaccharides are boiled in weak acetic acid to hydrolyse the aforementioned bond between the inner core and the hydrophobic lipid A. Intact lipopolysaccharides were placed in a spherical boiling flask with 1% acetic acid (2 mg LPS in 1ml acid) and placed in a preheated, circular dish containing hot sand. An open-top reflux condenser was attached, with running water flowing up and outward. The flask temperature was raised to 100°C and left to reflux for two hours on a hot-plate once a steady temperature was obtained. After this time, a cloudy precipitate is present in the solution. The solution was allowed to cool to 4°C and centrifuged at 5000G for 20 minutes. The lipid A component forms a small brown pellet, and the aqueous supernatant (containing the polysaccharide region only) was removed, the pellet washed, and placed in a separate tube for drying.

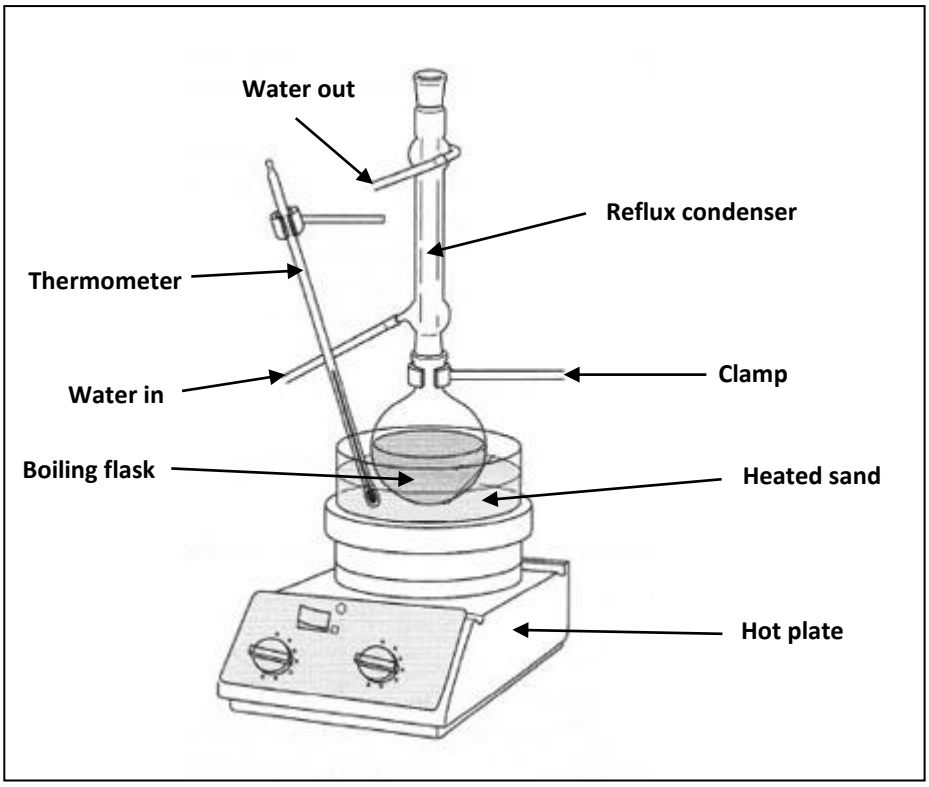


Figure 3-1 Schematic diagram of hydrolysis instrumental setup. The boiling flask contains the lipopolysaccharide and acetic acid which is raised to and maintained at 100°C boiling temperature. The reflux condenser ensures that the resultant vapour is cooled and re-collected in the flask. Image adapted from Mayo *et al*., 1994.

It has been well established that under mild acid conditions the Kdo residue that links the inner core to the lipid A moiety undergoes a conformational change and yields oligosaccharides with anhydro-Kdo on the reducing terminus rather than intact Kdo (Auzanneau *et al.,* 1991). Formation of anhydro-Kdo structures occurs when a phosphate group at the C4 position is β-eliminated (Danan *et al.,* 1982; Caroff *et al.,* 1987). This proceeds readily under these conditions because the C-4 substituent is β to the C-2 (anomeric) carbonyl carbon. The reaction is believed to give rise to olefinic Kdo derivatives, which rearrange to form anhydro ring structures – see figure 3-2.

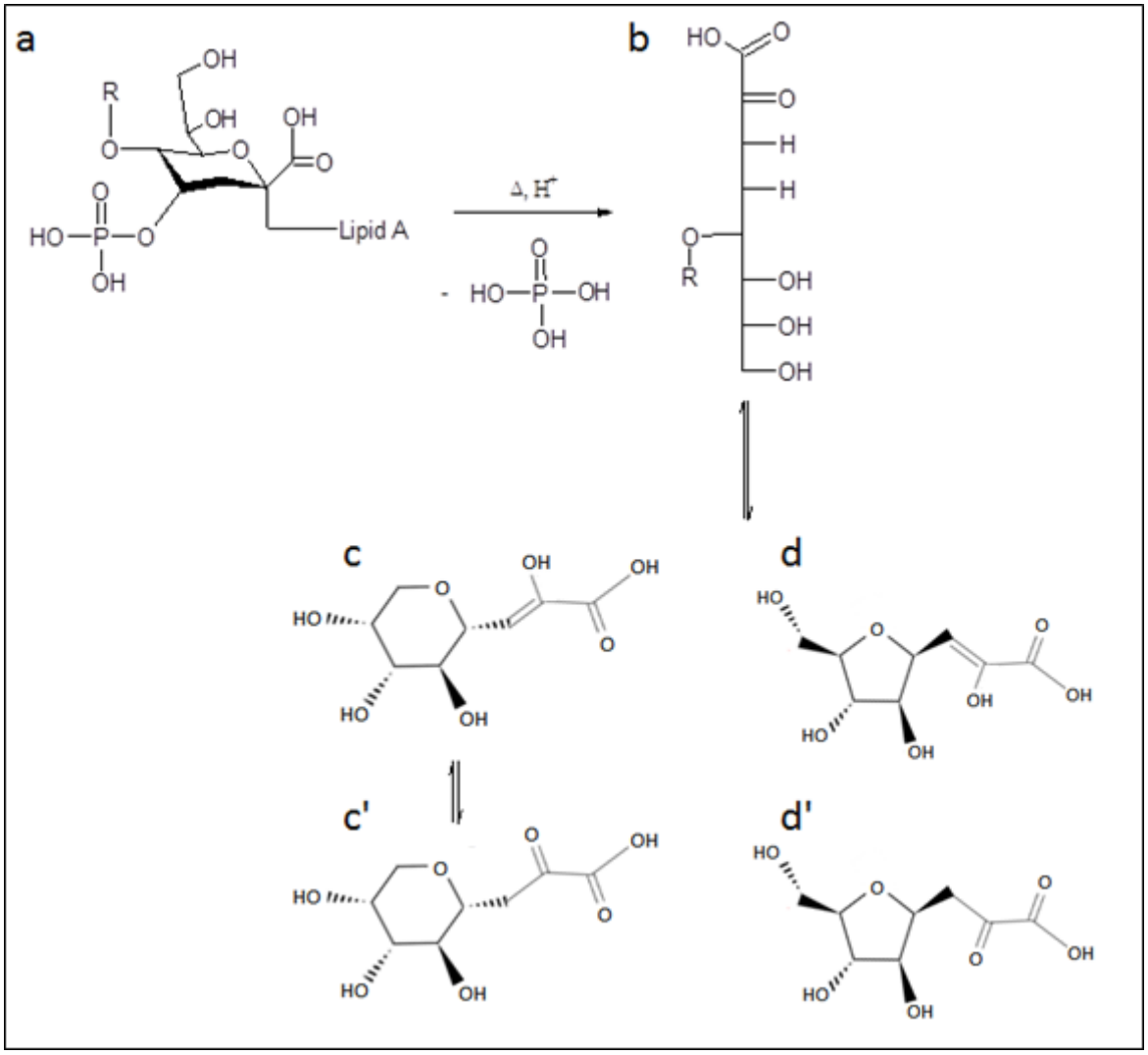


Figure 3-2 Formation of the structures formed during mild acid hydrolysis of 3-Deoxy-D-manno-oct-2-ulosonic acid (Kdo, a); D-arabino-3-en-2-ulonic acid open-chain (b), the diastereomeric forms of the 4,8-anhydro (c and c') and the 4,7-anhydro (d and d') derivatives. R represents the extended inner core of LPS. Image adapted from Sioud *et al*., 2010.

3.2.2 Lyophilisation

Both the supernatant and lipid A pellet from the hydrolysis were added to 15ml centrifuge tubes and stored at -80°C until frozen, when they were placed in a Savant Speedvac SPD111V with no heat, to dry-down the solvent. The vacuuming process was continued until only dry products and no liquid remained.

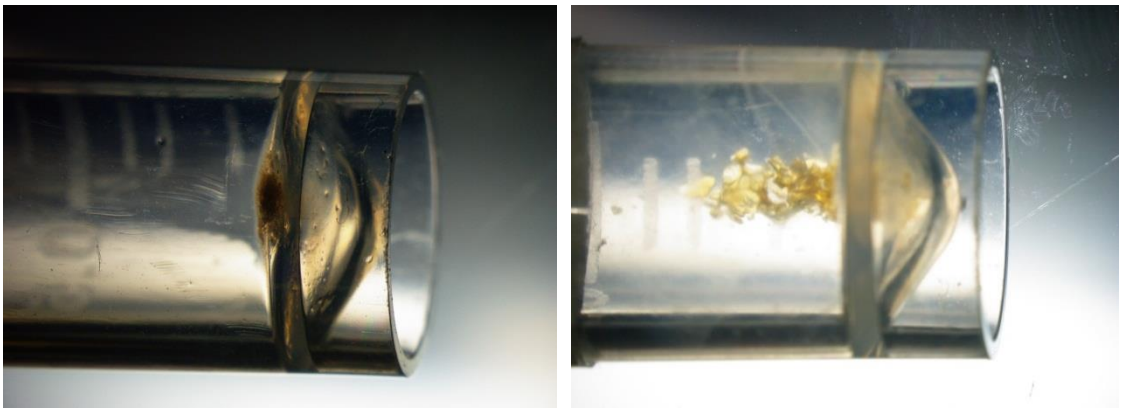


Figure 3-3 Photographs of tubes containing hydrolysed and lyophilised LPS. The left-hand tube contains the cleaved lipid A portion in the form of a small brown pellet, the right-hand tube contains the remaining polysaccharide region, with a crystalline consistency.

3.3 Crystallisation trials

Crystals of the recombinant head and neck fragment of human SP-D were grown in sitting drops, consisting of an equal volume (1µl) of protein at various mg/ml and precipitant buffer to screen for optimum conditions. See Appendix for tables of crystallisation screens and subsequent follow-up trials.

Some cocrystallisation trials were attempted. Tray CCS7 consisted of drops of protein preparation 1, dialysed in 10mM Tris, 10mM $CaCl_2$ and 140mM NaCl at pH 7.5, which was incubated with intact *Haemophilus influenzae* Eagan 4A lipopolysaccharide, and then re-concentrated to 8.16 mg/ml (protein).

CCS14 consisted of drops of protein preparation 3 at 6.52 mg/ml, which was incubated with 18.5mM hydrolysed *E. coli* B6 lipopolysaccharide and 11.1mM $CaCl_2$, incubated for 2 hours at room temperature.

Wells C1 and D1 of tray CCS16 consisted of drops of protein preparation 3 at 7.42 mg/ml incubated with 0.18mM hydrolysed *S. minnesota* lipopolysaccharide and 10mM $CaCl_2$.

Initial trials of conditions known to yield crystals (see Appendix table CCS2, 17/12/10) using protein preparation 3 were the first to show crystal growth - only after addition of calcium chloride eight weeks after the drops were laid down. Wells deemed particularly successful were reproduced, namely 16% PEG 6000 at pH6, 16% PEG 4000 at pH8, 16% PEG 10000 at pH6 and 16% PEG 8000 at pH7, and the period between drop lay-down and calcium addition was varied to assess optimal time.

3.3.1 Cryoprotection approach and ligand soaks

Soaking ligands into pre-existing crystals is often the method of choice to obtain protein-ligand complexes, owing to the ease of the technique (Hassell *et al.,* 2006). Protein crystals may however, be fragile and soaking time and the solvent used to dissolve a ligand may damage the crystal itself. Tests were performed to optimise the soaking time and ensure any conformational changes that may occur within the protein caused by ligand binding do not disrupt crystal contacts.

Unless otherwise stated, crystals were prepared for cryocooling by addition of successive 2µl aliquots of precipitant buffer containing 5%, 10% and 15% 2-Methyl-2,4-pentanediol (MPD), followed by two additions of 2µl of buffer containing 20% MPD, and an exchange of 10µl drop solution for 10µl of 20% MPD cryobuffer.

3.3.2 Data collection

All data sets presented in this thesis were collected on Macromolecular Crystallography (MX) beamlines at Diamond Light Source Ltd, Diamond House, Harwell Science and Innovation Campus, Fermi Avenue, Didcot, Oxfordshire OX11 0QX, United Kingdom.

3.3.2.1 Data processing

Data collected for crystals from trays CCS2C2, CCS2D6, CCS2D52 and from wells CCS12A33, CCS3D41a, CCS16A12 and CCS16A15 was automatically reduced by fast_dp and XIA2 immediately following data collection using the Automatic Software Pipeline (diamond.ac.uk). The automatically generated MTZ file was then reindexed and had a free-R flag added using Uniqueify.

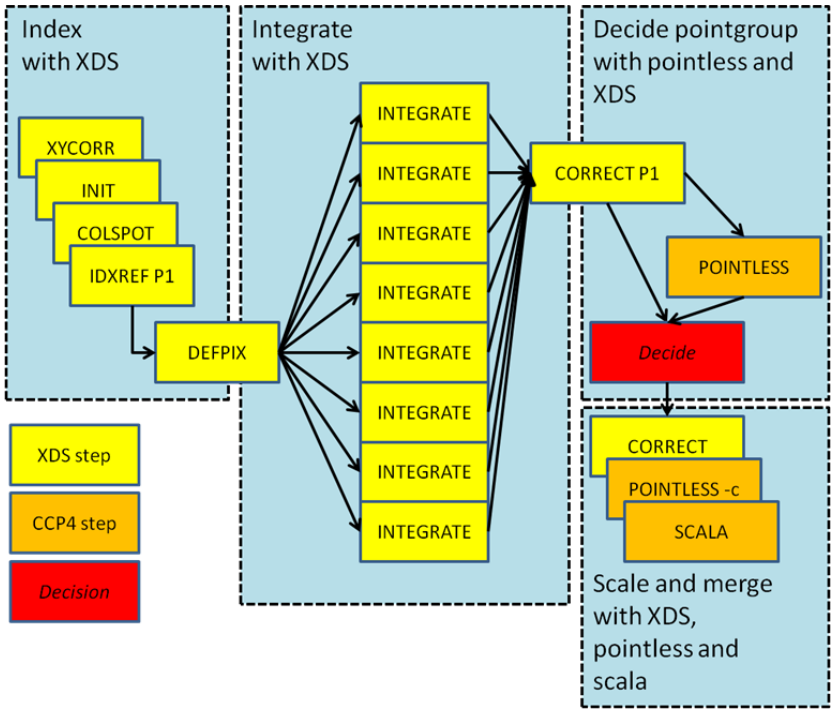


Figure 3-4 Flow-chart of fast data processing by fast_dp. Images are automatically indexed and integrated, the point groups is determined, then the images are scaled and merged to give an MTZ file containing the crystal reflection data (image taken from diamond.ac.uk).

Data processing and reduction for crystals from CCS9B3, CCS8C1, CCS3D41 and CCS3D42 was performed using the program MOSFLM. First, the diffraction spots were found using Find Spots;

indexed; cell refinement was performed to determine the cell parameters accurately using three separate segments at a range of different φ values unique and anomalous data; then integrated. CCP4 program SCALA was used to scale together multiple observations of reflections.

For crystal CCS16A11, integrated intensities were calculated using iMosflm. First, the diffraction spots were indexed; a strategy was calculated using two initial images, 90° apart in φ to evaluate unique and anomalous data (42.00-42.10° and 131.90-132° φ); cell refinement was performed to determine the cell parameters accurately using three separate segments at a range of different φ values (images 1-32, 450-481 900-932); and finally, all images were integrated.

After integration, the program Pointless was run to determine the point group, Aimless was run to scale together multiple observations of reflections, and CTruncate was run to calculate intensities. Using FreeRflag, a tag each for reflection in the MTZ file with a flag for cross-validation was added, corresponding to a 4.869% subset of reflections not to be used in the refinement.

3.3.2.2 Structure solution

Isomorphism was sufficient to allow the coordinates of a previously determined structure of a trimeric SP-D fragment (residues 208–355 in chains A, B and C) to be used as a starting model for the structure solution of all processed data sets. For all crystals (with the exception of those from tray CCS16), structure solution was carried out using the CCP4 program suite: SIGMAA was used to calculate Fourier coefficients from the calculated phases by combining phase probabilities from isomorphous phases from the rigid body refinement using a native protein structure without water molecules or calcium – completed using CNS of the CCP4 suite. Density modification and solvent flattening were performed using program DM. For crystal CCS16A11 only, a rigid body refinement was accomplished through CCP4 program REFMAC using a native protein structure without water molecules, at a resolution of 1.77 Å chosen to conform to reasonable signal-to-noise ratio. 20 cycles of refinement were performed automatically.

3.3.2.3 Electron density maps and refinement

For all data sets besides CCS16A11, fast Fourier transforms (FFT) were utilised to generate electron density maps, and model building was accomplished using O. No refinement was performed.

For CCS16A11 weighted difference maps were generated during rigid body refinement by REFMAC. Ligand was then modelled in JLigand using known structures of heptose monosaccharides and 4,7-anhydro Kdo from Sioud *et al.,* 2010. Dictionaries of torsion/bond distances and angles were calculated for the regularised ligand using Sketcher. The structure was refined once positions of amino acid side- and main-chain conformations had been fitted in the electron density using Coot, and repeated. Restrained refinement of the ligand-bound structure was carried out, using the aforementioned dictionaries. Water molecules were added using FindWaters in the Coot suite, based on density peaks that fit certain criteria: peaks above a median electron density of 1.8 rmsd and 2.4-3.2 Å distance to protein atoms. CheckWaters then removed molecules with a B-factor greater than 80 $Å^2$ and a map rmsd level less than 1.00 electron per $Å^3$. After addition of water molecules, resolution was sufficient to complete a final round of refinement. Refinement statistics are shown in table 4-6 in the Results chapter.

Chapter 4: Results of crystallisation trials of recombinant human SP-D

This chapter details the outcomes of crystallisation trials and refinements, and subsequent collection of successful data sets - along with the electron density maps and solved structure for crystal CCS16A11.

4.1 Crystal trials and crystal growth

Crystal growth was generally successful, with crystals of SP-D consistently showing a rectangular morphology with regular, well-defined edges. Protein preparation 3 which had additional endotoxin treatment, was by far the most successful preparation with regards to crystal growth. Crystal trays with protein preparations 1 and 2 did not yield any crystals, and cocrystallisation trials in trays CCS7 and CCS14, in addition to wells in tray CCS16 yielded no crystals, but formed cloudy precipitate (see Fig 4-1).

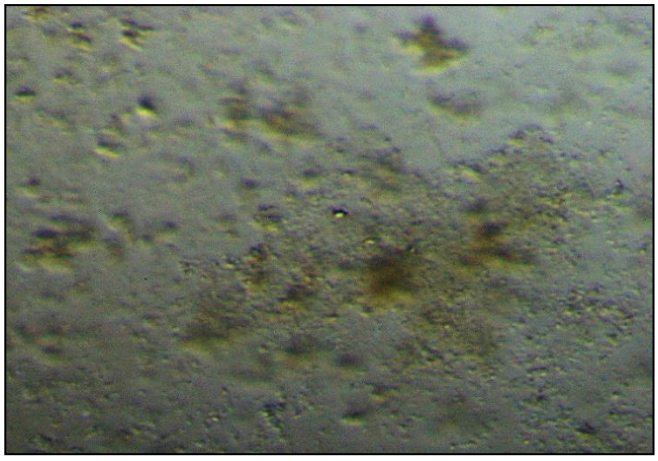


Figure 4-1 Well CCS7D2 showing typical formation of cloudy precipitate during cocrystallisation trials using protein preparation 1 which was dialysed in 10mM Tris, 10mM $CaCl_2$ and 140mM NaCl at pH 7.5, then incubated with intact *Haemophilus influenzae* Eagan 4A lipopolysaccharide and reconcentrated.

4.1.1 Crystallisation trials

Initial trials of conditions known to yield crystals (see Appendix table 2) using protein preparation 3 were the first to show crystal growth – but only after addition of calcium chloride eight weeks after the drops were laid down. Wells deemed particularly successful were reproduced, and the optimum period between drop assembly and calcium addition was demonstrated to be between 1 day and 2 weeks.

4.2 Ligand soaks

Cryoprotection soaks of pre-existing crystals using cryobuffers with ligand included were tested on crystals considered lower quality, to optimise soaking time and ensure the cryobuffer did not damage the crystal (see Appendix table 17). The following subchapters detail the lipopolysaccharides used in ligand soaks, in addition to the crystals that were soaked.

4.2.1 Intact *Haemophilus influenzae* 4A

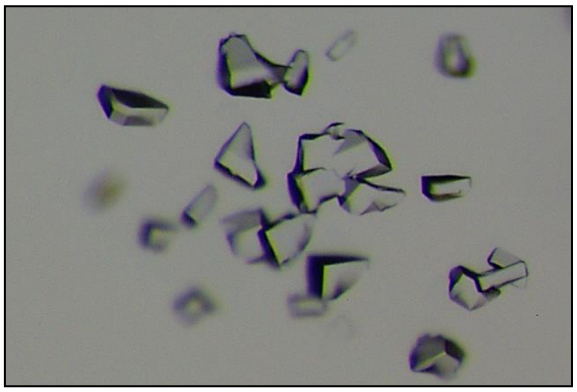


Figure 4-2 Crystals from well CCS2A3.

A trial of soaking intact 4A ligand was attempted by addition of 1µl 6.7mM ligand directly to the drop (CCS2B5, Fig 4-3), which resulted in a brown, murky precipitate forming in the centre. Crystals gradually degraded from the inside-out and by 24 hours, all had dissolved.

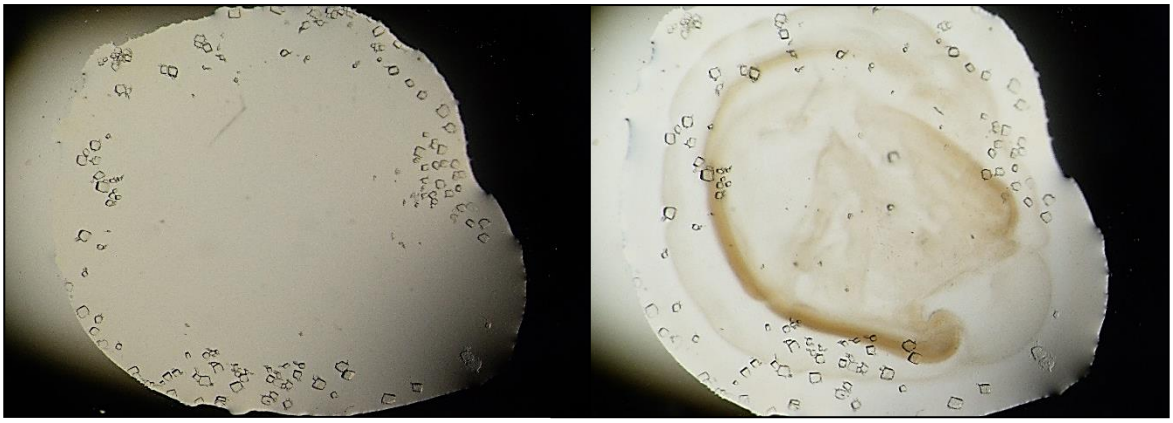


Figure 4-3 Crystals from well CCS2B5 test well before and immediately following addition of intact *H. influenzae* Eagan 4A LPS to drop. There was an instant formation of brown precipitate and gradual degradation of crystal integrity.

In well CCS2A3, two crystals were cryoprotected and frozen without ligand, then 12µl 6.7mM Eagan 4A LPS was added, and two further crystals were frozen just over an hour later. Two weeks later the well was re-exchanged and an additional 6µl 6.7mM ligand was added. Three more crystals were subsequently frozen within 30 minutes. In total, eight crystals were taken from well CCS2A3 and data collected for three (CCS2A34, A35 and A37); the first two of which had data collected on beamline I03 while A37 was tested on beamline I04-1 on a different date.

4.2.2 *E. coli* B4 smooth

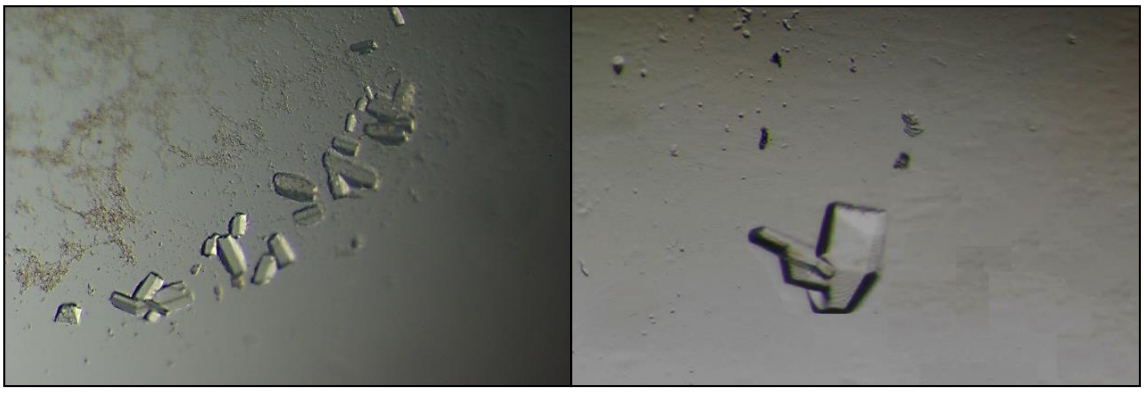


Figure 4-4 Crystals from wells CCS2C2 (left) and CCS2D6 (right).

Tests for soaking with 30mM hydrolysed *E. coli* B4 smooth LPS resulted in quick degradation of the crystals, and in one case crystals dissolved immediately. During preparation for freezing using CCS2C2 and CCS2D6 however, little effect was seen and wells were able to be exchanged after 3 hours 45 minutes for CCS2C2 and 56 minutes for CCS2D6. Five crystals were frozen from the former well, and two from the latter - all of which had data collected on beamline I03.

4.2.3 *E. coli* B6 smooth

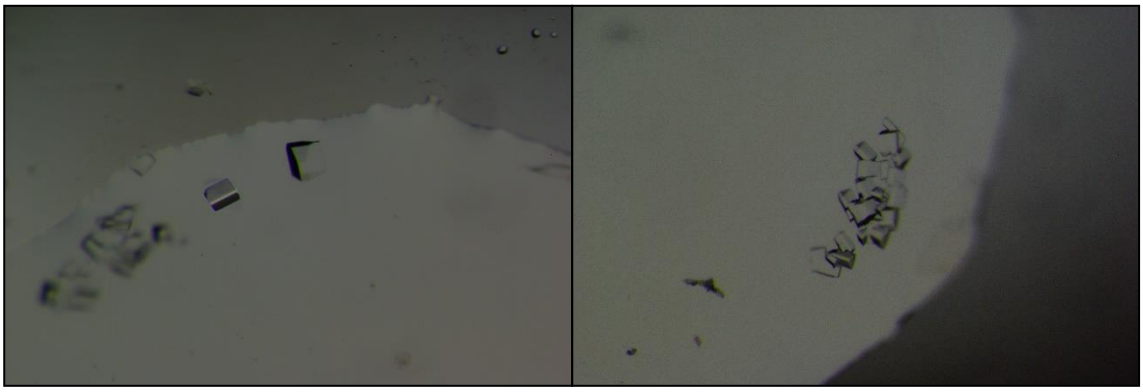


Figure 4-5 Crystals from wells CCS9B3 (left) and CCS9A4 (right)

Crystals from wells CCS9B3 and CCS9A4 were cryoprotected and soaked with 30mM hydrolysed lipopolysaccharide from *E. coli* B6 and exchanged immediately. Three crystals from CCS9B3 were collected and frozen roughly 30 minutes post-exchange, and a single crystal from CCS9A4 was frozen after 10 minutes. The three crystals from CCS9B3 had data collected at beamline I04.

4.2.4 *E. coli* F583 rough

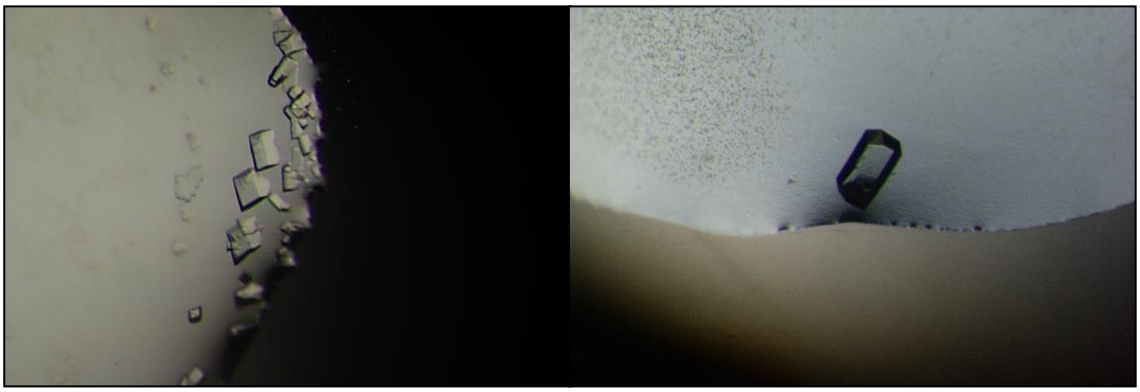


Figure 4-6 Crystals from well CCS12D5 (left) and CCS12A3 (right)

Test wells soaked with 10mM *E. coli* F583 demonstrated that crystals were immediately damaged by introduction of the ligand, and within 5 minutes the drop became too cloudy to accurately see crystals at all. Therefore, soaking time was limited to as short a period of time as possible: six crystals were exchanged and frozen within 14 minutes from well CCS12D5, and three crystals were exchanged and frozen within three minutes from well CCS12A3. Data was collected for crystals CCS12D51, D51 and D55; and CCS12A32 on beamline I02.

Figure 4-7 Dissolved crystals from test well soaked with hydrolysed *E. coli* F583 polysaccharide.

4.2.5 *E. coli* J5 rough

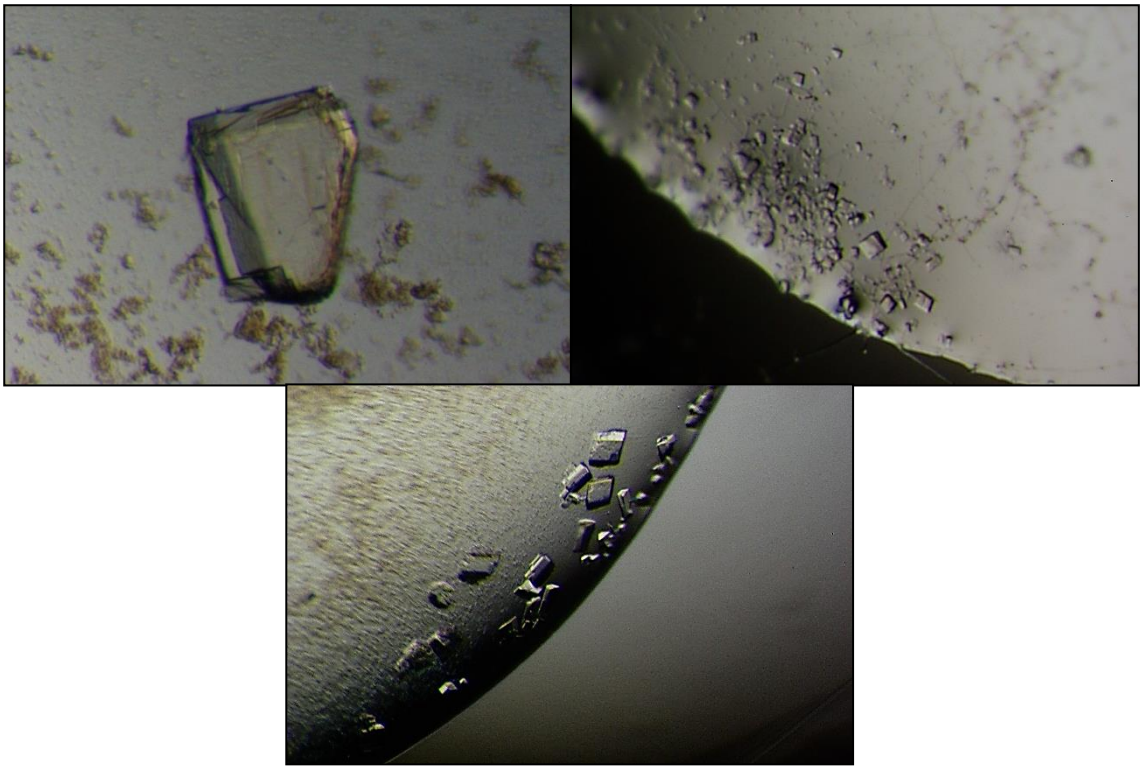


Figure 4-8 Crystals from wells CCS8C1 (left), CCS3D4 (right) and CCS13B6 (bottom).

Crystals from wells CCS8C1 (initially assumed to be a chunk of glass) and CCS3D4 were not affected by cryobuffers containing 30mM hydrolysed *E. coli* J5 lipopolysaccharide and as such were able to withstand a wait of 44 minutes to over an hour between exchange and freezing. A single crystal was frozen from well CCS8C1, and three taken from CCS3D4. Of these four crystals, four underwent data collection on beamline I04 (two data sets were collected for crystal CCS3D41); crystal CCS3D43 showed no diffraction. Crystals from CCS13B6 were subjected to successive aliquots of 5-20% cryobuffer, then were exchanged after 40 minutes and four crystals were frozen within 9 minutes. Data was collected for crystals CCS8C1, CCS3D41 and D42 on beamline I04. Data sets from well CCS13B6 were not processed.

4.2.6 *Salmonella minnesota* R7 rough

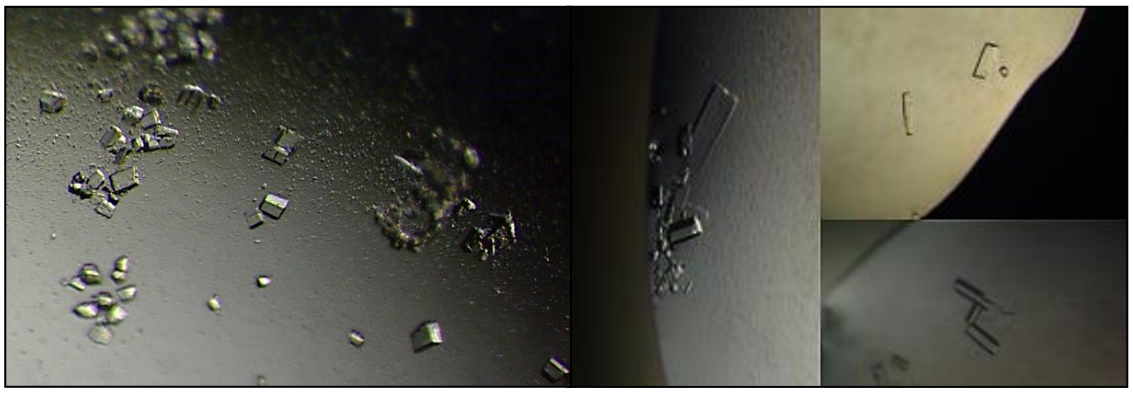


Figure 4-9 Crystals from wells CCS2C6 (right) and CCS16A1 (left).

Test wells which were soaked with cryobuffers containing 30mM hydrolysed *S. minnesota* R7 polysaccharide suggested that crystals degraded very quickly once soaked. Nine crystals were taken from well CCS2C6, which were frozen as soon as possible after exchange due to immediate degradation. Unfortunately, none produced diffraction spots and as such no data sets were collected. For well CCS16A1, following addition of the 5%, 10% and 15% MPD + R7 polysaccharide cryobuffers, crystals again appeared to be degrading and three crystals were removed from the well and frozen before addition of 20% cryobuffer to the well. 1µl of the 20% cryobuffer was then added to the well, followed by exchanging 8µl of the well with the 15% cryobuffer with added calcium (1µl 100mM $CaCl_2$). The crystals appeared to show no further degradation, and six more crystals were frozen between 7 minutes and 1 hour 18 minutes after the buffer exchange. Of the nine crystals frozen from well CCS16A1 five had data collected - a diffraction image from crystal CCS16A11 (the first crystal frozen from this well) is shown below.

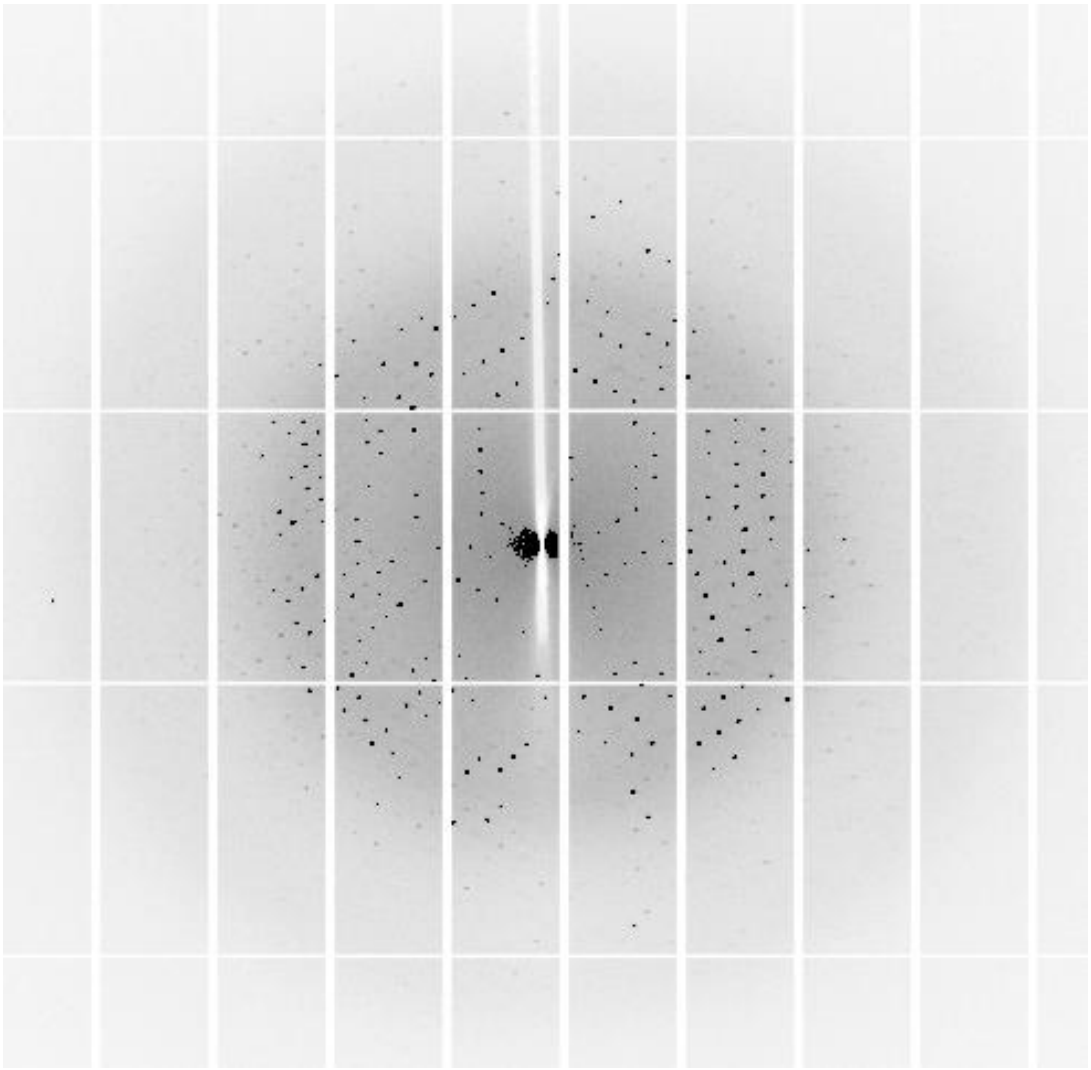


Figure 4-10 Diffraction pattern from a crystal CCS16A11. Reflections form a regular pattern but intensity (spot darkness) varies, and fade toward the outer regions of the image - corresponding to the data set's resolution of roughly 1.6 Å. Image collected on MX beamline I03 at Diamond Light Source and is one of 1,300 images taken for this crystal.

4.3 Data collection

All data sets presented in this thesis were collected on Macromolecular Crystallography (MX) beamlines at Diamond Light Source Ltd, Diamond House, Harwell Science and Innovation Campus, Fermi Avenue, Didcot, Oxfordshire OX11 0QX, United Kingdom. Details of all collections are shown in the table below, including the resolution at which data sets were collected - a diffraction pattern collected for crystal CCS16A11 is shown in Fig. 4-10. Details of crystals which were tested or soaked can be found in appendix Table 17.

The following table details the cryoprotection procedure and whether the crystal yielded diffraction data.

Table 4-1 Exposed crystal cryoprotection procedure, ligand concentration and subsequent exposures performed, with diffraction resolution, DLS beamline and date of collection where applicable. Cryo - cryobuffer; Snap - snapshot taken to assess diffraction quality; DC - data collection. See Appendix for full details of crystallisation conditions. Data for crystals highlighted in yellow were processed.

Crystal	Cryoprotection procedure	Ligand concentration	Result	
CCS2 A3 1 CCS2 A3 2	Exchanged and frozen, native 4/5/11	-	Snap 1.75 Å Snap 1.8 Å	7/5/11 I03
CCS2 A3 3 CCS2 A3 4 CCS2 A3 5	Exchanged and frozen + 12µl 6.7mM Eagan 4A LPS 4/5/11	3.3mM	Snap 9 Å Collected 1.8 Å Collected 1.8 Å	7/5/11 I03
CCS2 A3 6 CCS2 A3 7 CCS2 A3 8	Reexchanged + 6µl 6.7mM Eagan 4A LPS 18/5/11	2.86mM	Snap 2.5 Å DC 1.5 Å Snap 2.86 Å	20/5/11 I04-1
CCS2 B1 1 CCS2 B1 2 CCS2 B1 3 CCS2 B1 4	Exchanged and frozen, native 1/7/11	-	No diffraction	4/7/11 I04-1
CCS3 C1 1 CCS3 C1 2	Exchanged and frozen, native 7/3/12	-	DC 2.5 Å Snap 4.2 Å	6/5/12 I04
CCS2 C2 1 CCS2 C2 2 CCS2 C2 3 CCS2 C2 4 CCS2 C2 5	5-20% MPD cryo containing 30mM *E. coli* B4 PS. Exchange after 3h 45m 15/3/12	30mM	DC 2.4 Å DC 2 Å DC 2.4 Å DC 2.5 Å DC 2 Å	17/3/12 I03
CCS2 D6 1 CCS2 D6 2	5-20% MPD cryo containing 30mM *E. coli* B4 PS. Exchanged after 56m 15/3/12	30mM	DC 1.6 Å DC 1.7 Å	17/3/12 I03
CCS9 B3 1 CCS9 B3 2 CCS9 B3 3	5-20% MPD cryo containing 30mM *E. coli* B6 PS. Exchanged immediately. 4/5/12	30mM	DC 1.9 Å DC 2.0 Å DC 1.9 Å	6/5/12 I04

Table 4-2 Exposed crystal cryoprotection procedure, ligand concentration and subsequent exposures performed, with diffraction resolution, DLS beamline and date of collection where applicable. Cryo - cryobuffer; Snap - snapshot taken to assess diffraction quality; DC - data collection. See Appendix for full details of crystallisation conditions. Data for crystals highlighted in yellow were processed.

CCS9 A4 1	5-20% MPD cryo containing 30mM *E. coli* B6 PS. Exchanged immediately. 4/5/12	30mM	DC 2.0 Å	6/5/12 I04
CCS3 D4 1 CCS3 D41a CCS3 D4 2 CCS3 D4 3	5-20% MPD cryo containing 30mM *E. coli* J5 PS. Exchanged after 40 minutes. 17/7/12	30mM	DC 3 Å DC 2.5 Å DC 2.5 Å No diffraction	22/7/12 I04
CCS8 C 1	Re-exchanged with 20% MPD cryo containing 30mM *E. coli* J5 PS (10µl out, 10µl in) 17/7/12	30mM	DC 1.5 Å	22/7/12 I04
CCS13 B6 1 CCS13 B6 2 CCS13 B6 3 CCS13 B6 4	5-20% MPD cryo containing 30mM *E. coli* J5 PS. Exchanged after 40 minutes. 6/8/12	30mM	Not tested DC 1.9 Å DC 2.14 Å DC 1.7 Å	9/8/12 I03
CCS12 D5 1 CCS12 D5 2 CCS12 D5 3 CCS12 D5 4 CCS12 D5 5 CCS12 D5 6	5-20% MPD cryo containing 10mM *E. coli* F583. Exchanged immediately. 10/10/12	10mM	DC 2.4-3 Å DC 3.18 Å Not tested Not tested DC 3.0 Å Not tested	14/10/12 I02
CCS12 A3 1 CCS12 A3 2 CCS12 A3 3	5-20% MPD cryo containing 10mM *E. coli* F583. Exchanged immediately. 10/10/12	10mM	DC 1.4 Å Not tested DC 2-3 Å	14/10/12 I02
CCS2 C6 1 CCS2 C6 2 CCS2 C6 3 CCS2 C6 4 CCS2 C6 5 CCS2 C6 6 CCS2 C6 7 CCS2 C6 8 CCS2 C6 9	5-20% MPD cryo containing 22mM *S. minnesota* R7. Exchanged immediately. 15/3/13	22mM	No diffraction	20/3/13 I04
CCS16 A5 1	5-20% MPD cryo containing 22mM *S. minnesota* R7. Exchanged immediately. 15/3/13	22mM	No diffraction	20/3/13 I04
CCS16 A1 1 CCS16 A1 2 CCS16 A1 3	5-15% MPD cryo containing 22mM *S. minnesota* R7. Crystals degrading – exchanged before 20% aliquots	22mM	DC 1.63 Å DC 2.13 Å Not tested	19/5/13 I03
CCS16 A1 4 CCS16 A1 5 CCS16 A1 6 CCS16 A1 7 CCS16 A1 8 CCS16 A1 9	As above. 1µl 20% cryo added, 20µl removed from drop. Total volume of remaining 15% cryo (plus 1µl 100mM $CaCl_2$) added	17.6mM	Not tested DC 2.4 Å Not tested Not tested DC 1.7 Å DC 2.0 Å	19/5/13 I03

4.4 Data processing

As mentioned previously, data collected for crystals from tray CCS2C and from wells CCS12A3, CCS3D4, and CCS16A1 was automatically reduced by fast_dp and XIA2 immediately following data collection using the Automatic Software Pipeline (diamond.ac.uk). The generated MTZ file containing reflection data, symmetry information and cell dimensions was then reindexed and had a free-R flag added using Uniqueify.

Data for crystal data sets CCS9B3, CCS8C1, CCS3D41 and CCS3D42 were processed and reduced using MOSFLM. Firstly, the diffraction spots were found using Find Spots; they were then indexed and a cell refinement was performed to determine the cell parameters accurately. The data was then integrated using the CCP4 program SCALA, which was used to scale together multiple observations of reflections. These data sets however, showed no ligand in the binding site, and as such are not discussed further in this thesis. The following table details the data collection measurements and processing method of each individual data set collected:

Table 4-3 Measurements of collected data sets. Data sets with superscript symbols saw a drop of diffraction quality during collection. The term 'Fast' in the processing method column refers to data that underwent fast_dp and XIA2 immediately following data collection using the Automatic Software Pipeline; 'Full' refers to data sets that were reindexed and integrated through MOSFLM, and scaled through SCALA.

Crystal	° between images	Number of images	% transmission	Processing method	Resolution (Å)
Intact *H. influenzae* LPS					
CCS2 A3 4	0.1	2000	43	Fast	1.8
CCS2 A3 5	0.1	2000	43	Fast	1.85
CCS2 A3 7	0.2	900	100	Fast	1.5
Hydrolysed *E. coli* B4 PS					
CCS2 C2 1	1	200*	200	Fast	2.4
CCS2 C2 2	0.2	900	100	Fast	2.0
CCS2 C2 3	0.2	1000	50	Fast	2.4
CCS2 C2 4	0.2	900	50	Fast	2.5
CCS2 C2 5	0.2	900	50	Fast	2.0
CCS2 D6 1	0.1	2000	50	Fast	1.6
CCS2 D6 2	0.1	2000	50	Fast	1.7
Hydrolysed *E. coli* B6 PS					
CCS9 B3 1	1	200	100	Full	1.9
CCS9 B3 2	1	200	100	Full	2.0
CCS9 B3 3	1	200	100	Full	1.9
Hydrolysed *E. coli* J5 PS					
CCS3 D4 1	1	200	100	Full	3.0
CCS3 D4 1a	0.5	200	100	Fast	2.5
CCS3 D4 2	1	200	100	Full	2.5
CCS8 C 1	0.5	400	50	Full	1.5
Hydrolysed *E. coli* F583 PS					
CCS12 D5 2	0.1	2000	50	Fast	3.18
CCS12 A3 3	0.1	2000	60	Fast	2.0

Table 4-4 Measurements of collected data sets. Data sets with superscript symbols saw a drop of diffraction quality during collection. The term 'Fast' in the processing method column refers to data that underwent fast_dp and XIA2 immediately following data collection using the Automatic Software Pipeline; 'Full' refers to data sets that were reindexed and integrated through MOSFLM, and scaled through SCALA.

Hydrolysed *S. minnesota* R7 PS					
CCS16 A1 1	0.1	1310	40	Full	1.63
CCS16 A1 2	0.1	1800	60	Fast	2.13
CCS16 A1 5	0.1	1800	50	Fast	2.4

[*] Quality drop after image 80, [+] after image 600, [#] after image 200.

Table 4-5 Results of data sets that underwent processing to generate preliminary electron density maps that showed no ligand in the binding site

Crystal	Soaked with	Processing details	Result
CCS2 A3 3 **CCS2 A3 4** **CCS2 A3 7**	*Haemophilus influenzae* 4A	Automatically reduced by fast_dp (viewed with O).	No ligand density was seen in the binding site
CCS2 C2 1 **CCS2 C2 2** **CCS2 C2 3** **CCS2 C2 4** **CCS2 C2 5** **CCS2 D6 1** **CCS2 D6 2**	*E. coli* B4 smooth	Automatically reduced by fast_dp (viewed with O).	No ligand density was seen in the binding site
CCS9 B3 1 **CCS9 B3 2** **CCS9 B3 3**	*E. coli* B6 smooth	Processed through MOSFLM and SCALA (viewed with O).	No ligand density was seen in the binding site
CCS3 D4 1 **CCS3 D41a** **CCS3 D4 2** **CCS8 C1**	*E. coli* J5 rough	for CCS8C1, CCS3D41 and D42 was processed through MOSFLM and SCALA; data for CCS3D41a was automatically reduced by fast_dp; data sets from well CCS13B6 were not processed	Large areas of electron density were seen in the ligand-binding site - however, the definition was ambiguous and no definite carbohydrate could be discerned.
CCS12 D5 1 **CCS12 D5 2** **CCS12 D5 5**	*E. coli* F583	Automatically reduced by fast_dp (viewed with O).	No ligand density was seen in the binding site

The following figure, Fig. 4-11, shows an electron density map of the ligand binding site with a marked absence of bound lipopolysaccharide, and is representative of each of the data sets in Table 4.5.

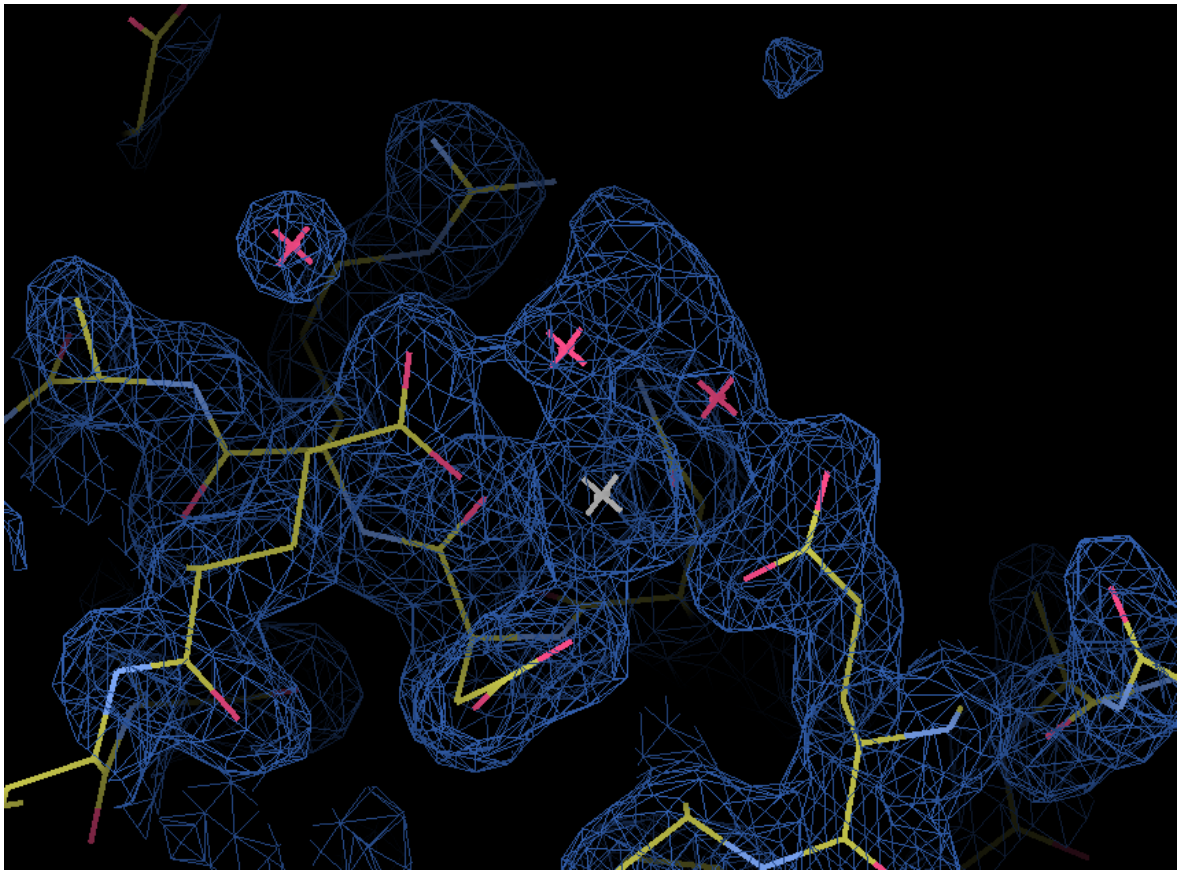


Figure 4-11 A 2Fo-Fc difference map showing no ligand in the CRD. The central white cross is calcium 1 - the site of carbohydrate binding, while the pink crosses are water molecules whose positions are otherwise occupied by hydroxyl pairs of a ligand. Image generated using Coot.

For crystal CCS16A11 (soaked with hydrolysed *S. minnesota* R7 PS), integrated intensities were calculated using iMosflm. First, the diffraction spots were indexed; a strategy was calculated using two initial images, 90° apart in φ to evaluate unique and anomalous data (42.00-42.10° and 131.90-132° φ); cell refinement was performed to determine the cell parameters accurately using three separate segments at a range of different φ values (images 1-32, 450-481 900-932); and finally, all images were integrated.

After integration, the program Pointless was run to determine the true Laue symmetry, Aimless was run to scale together multiple observations of reflections, and CTruncate was run to calculate cumulative intensity distributions and a Wilson plot as well as to perform any anisotropy correction.

Using FreeRflag, a tag each for reflection in the MTZ file with a flag for cross-validation was added, corresponding to a 4.869% subset of reflections not to be used in the refinement.

Table 4-6 Data collection and merging statistics for crystal CCS16A11.

Wavelength (Å)	0.97625
Temperature (K)	100
Spacegroup	$P2_1$
a (Å)	55.32
b (Å)	108.28
c (Å)	55.68
α (°) = γ (°)	90.00
β (°)	91.75
Max resolution (Å)	1.7
Observations	224,221
Unique reflections	99,332
Completeness (%)	94.1
R_{merge}	0.085
B-factor (protein)	22.7
B-factor (ligand)	25.5
B-factor (water)	33.3
Mean I/σI	4.8
Highest resolution bin	
Highest resolution bin (Å)	1.77-1.82
Completeness (%)	91.6
R_{merge}	0.304
Mean I/σI	2.6

4.5 Structure solution and maps

As the data for crystal CCS16A11 gave the most promising preliminary electron density map, the structure solution for this set only is described here. Structure solution and model building was performed as previously described in Chapter 3. Isomorphism was sufficient to allow the coordinates of a previously determined structure of a trimeric SP-D fragment (PDB accession code 1PW9) to be used as a starting model for the structure solution of all processed data sets.

A rigid body refinement was accomplished through CCP4 program REFMAC using the native protein structure without water molecules (PDB accession code 1PW9), at a resolution of 1.77Å (see Table 4-6). 20 cycles of refinement were performed automatically and reduced the R-factor from 32.2% to 18.5% and the R-free (equivalent to R-factor for a 4.87% subset of reflections not used in the refinement) from 32.2% to 22.1%. The improved phases gave a significantly clearer position of the anKdo tail group - and defined its orientation towards Arg343 (see figures 4-16 and 4-17).

B-factors, calculated using CCP4i program BAVERAGE, were 23.6 for protein in the A chain, 21.1 for the B chain, 23.4 in the C chain, 25.5 for the ligand (averaged over both ligand-binding chains B and C), and 33.3 for water molecules.

4.5.1 CCS16A11 - *S. minnesota* R7-bound structure

The overall structure of the recombinant head and neck fragment of human SP-D reveals a trimeric structure of three C-terminal globular head domains, linked by an extended α-helical coiled coil neck region, with an N-terminal collagen domain that is not visible in the electron density. The structure contains residues 205-355 in chains A and B, and 206-355 in chain C.

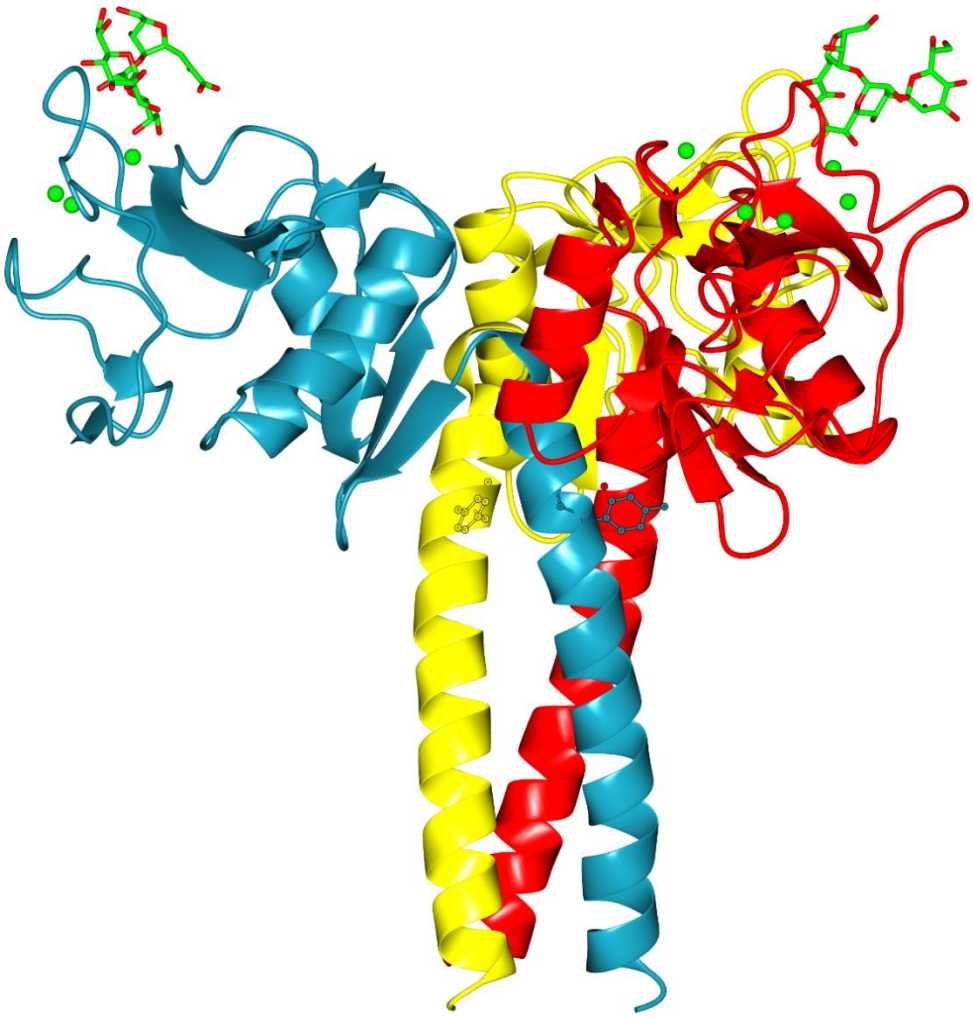


Figure 4-12 Crystal structure of *S. minnesota* R7 polysaccharide-bound SP-D showing three calcium ions per CRD in green, and the asymmetric tyrosine positioned into the central pore. Chain A is shown in yellow, chain B in cyan and chain C in red. Ligand is seen in the B and C chains. Image generated using CCP4MG.

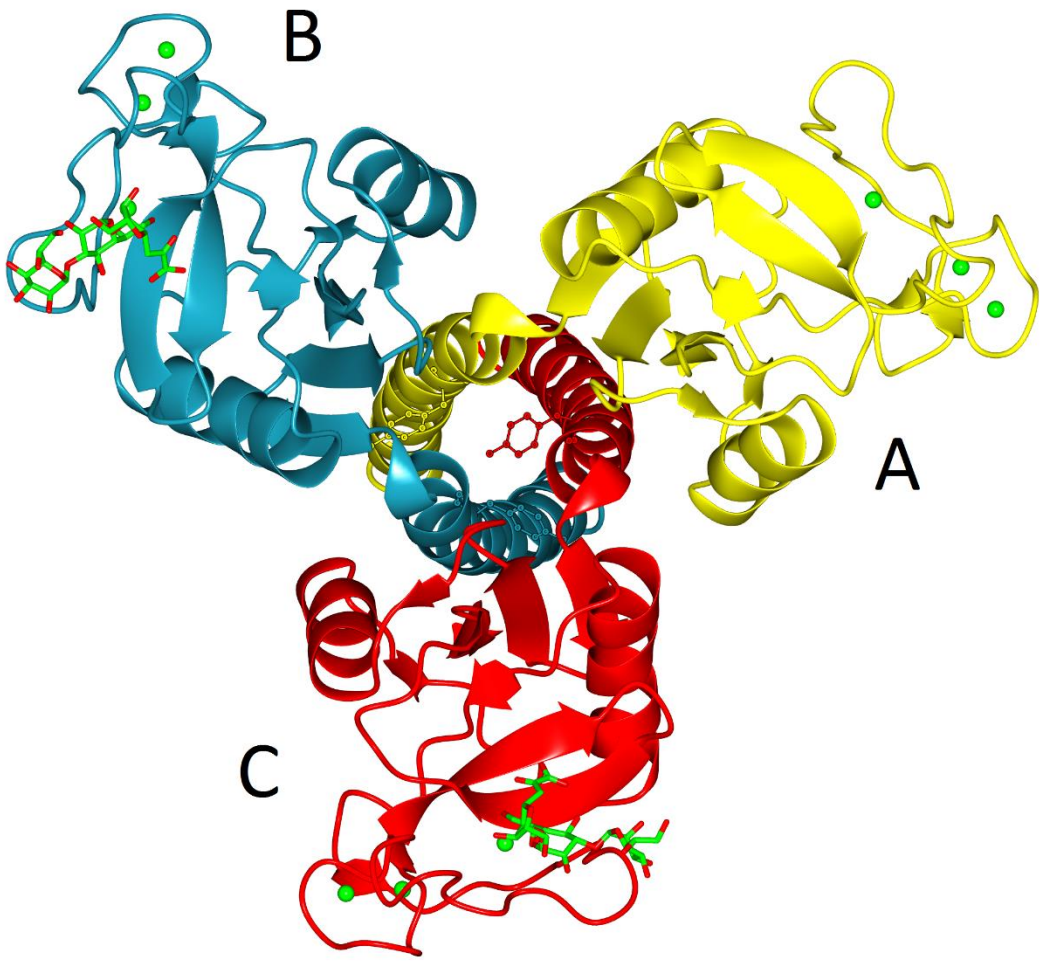


Figure 4-13 Crystal structure of *S. minnesota* R7 polysaccharide-bound SP-D in ribbon form, showing three calcium ions per CRD in green, and the asymmetric tyrosine positioned into the central pore. Ligand is seen in the B and C chains. Image generated using CCP4MG.

Similar to structures elucidated by Shrive *et al*., 2003, the electron density suggested alternative conformations of the sidechains of Cys353 and Cys345 which formed disulphide bridges with Cys261 and Cys331 respectively. These features are consistent with radiation damage to thiol groups of cysteine residues, even at cryogenic temperatures, using moderate-intensity synchrotron radiation (Weik *et al.,* 2000).

4.5.1.1 Calcium ions

Three calcium ions per CRD were clearly visible in the electron density: calcium 1 in the ligand binding vicinity which is coordinated by the sidechains of Glu321, Asn323, Glu329, Asn341, Asp342 and hydroxyl groups 6 and 7 of the bound ligand. Calcium 2 is coordinated by the sidechains of Asp297, Glu301, Asp324 and Asp330 as well as the main-chain carbonyl of Glu329. Calcium 3 is coordinated by the sidechains of Glu301, Asp330 and Asp330 as well as four water molecules. The distances between calcium ions and coordinating amino acid residues are given in the table below:

Table 4-7 Coordination of calcium ions, water molecules or ligand residues by atoms of amino acid sidechains in the CRD of all three chains. Distances in Å

Calcium	Atom	A	B	C
Ca1	Glu321OE1	2.56	2.57	2.66
	Asn323OD1	2.42	2.45	2.36
	Glu329OE1	2.46	2.41	2.30
	Asn341OD1	2.39	2.39	2.45
	Asp342OD1	2.31	2.30	2.26
	Water/ligand	W388 2.53	O6 2.51	O6 2.50
	Water/ligand	W389 2.37	O7 2.51	O7 2.52
Ca2	Asp297OD1	2.70	2.61	2.71
	Asp297OD2	2.71	2.46	2.41
	Glu301OE1	2.49	2.55	2.52
	Glu301OE2	2.54	2.59	2.51
	Asp324OD1	2.39	2.55	2.59
	Glu329O	2.45	2.32	2.38
	Asp330OD1	2.26	2.39	2.39
Ca3	Glu301OE1	2.25	2.39	2.33
	Asp330OD1	2.49	2.52	2.52
	Asp330OD2	2.41	2.47	2.47
	Water	W386 2.29	W31 2.37	W21 2.37
	Water	W104 2.36	W58 2.40	W46 2.31
	Water	W87 2.37	W73 2.10	W13 2.24
	Water	W53 2.23	W94 2.24	W385 2.38

4.5.1.2 The asymmetric tyrosine 228 and the neck-CRD interaction

There is a marked asymmetry in the conformation of tyrosine residues at position 228 in the neck region. The sidechain of Tyr228 in the C chain is shifted into the centre of the coiled-coil, with the remaining two remaining directed outward (see Fig 4-13).

The fourth calcium ion, Ca4, in the central pore (previously demonstrated in the maltose-bound structure elucidated by Shrive *et al.,* 2003) is absent in this structure, and as such two of the three glutamic acid residues at 232 interact with each other as well as with Lsy246 in the same chain, while the third (in the B chain) adopts a conformation whereby it interacts with the asymmetric tyrosine 228 in the central pore.

4.5.1.3 The ligand binding site

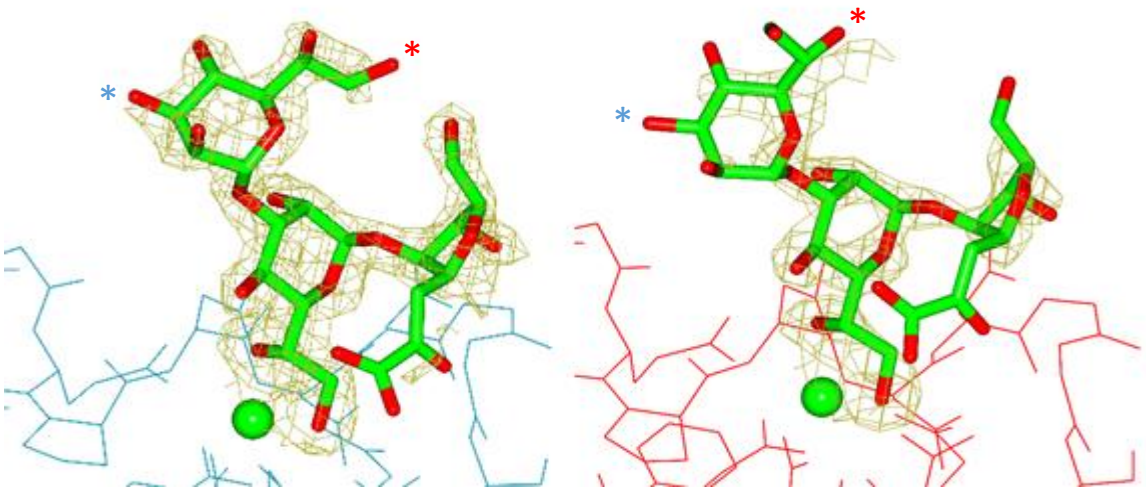


Figure 4-14 Fo-Fc difference map of bound *S. minnesota* R7 polysaccharide in chains B (left, blue) and C (right, red), calculated prior to adding ligand or water molecules into the model. The maps were contoured at 2.5σ. HepIII, which was not visible in the map, is linked to HepII via the O7 group, labelled with the red asterisk. The outer core region of LPS that is not present in this strain is connected via the O3 group of HepII, labelled with the blue asterisk. Image generated using CCP4MG.

Density corresponding to the ligand was well defined in the B chain, and to a lesser extent in the C chain, and showed the inner most heptose (HepI) binding to Ca1 by the 6 and 7 hydroxyl groups. The positions of the glycerol OH groups of the bound heptose (HepII) in the complex were similar to those of the equatorial 3 and 4 hydroxyl groups in previous ligand structures such as maltose and mannobiose (see Fig 1-20, replicated below).

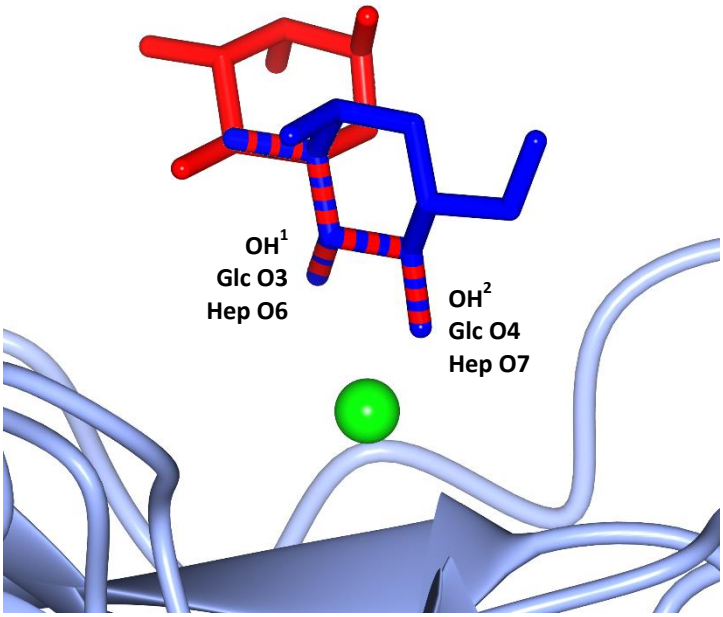


Figure 1-20 (replicated) Binding of glucose via vicinal hydroxyl groups O3 and O4 (Glc, blue) and heptose via hydroxyl groups of the glycerol side chain O6 and O7 (Hep, red). Bonds shown in alternate colours share the same coordinates in binding. OH^1 can be the O2 of galactose, O3 of α1-2mannobiose and α1-4mannobiose (in chain A only), O4 of α1-4mannobiose (in chains B and C), O4 of inositol-1-P (in chains B and C) and the O3 of maltose. OH^2 can be the O1 of galactose, O4 of α1-2mannobiose and α1-4mannobiose (in chain A only), O3 of α1-4mannobiose (in chains B and C), O5 of inositol-1-P (in chains B and C) and the O4 of maltose. Binding of myoinsoitol utilises the O1 and O6 hydroxyl groups which can take positions OH^1 and OH^2 reversibly. Image generated using CCP4MG using PDB accession codes 1PWB and 2RIB.

Evidence of the orientation and positions of the two adjacent carbohydrate residues HepII and Kdo was also clear, but density was lacking for the anhydro-Kdo (anKdo) tail group in both chains - and only weak, broken density was observed round hydroxyl groups 3, 4 and 6 of the second

heptose in chain C. HepIII is linked through the O7 hydroxyl group of HepII, but is not visible in the electron density.

The structure of the ligand was apparent however, and comprised two α1-3 linked heptose residues - the first of which linked to an anhydro, five-membered (rather than six-membered) Kdo residue known to form during mild acid hydrolysis.

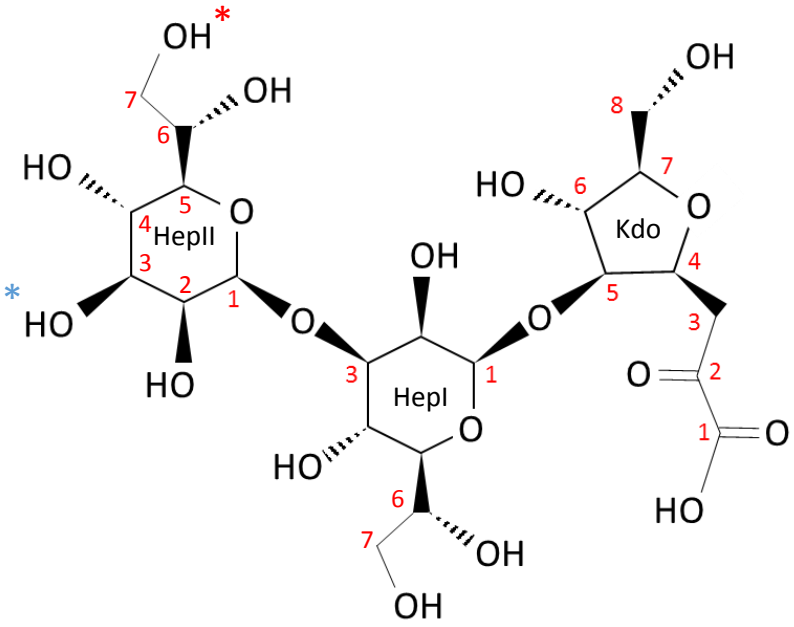


Figure 4-15 Structure of the polysaccharide from hydrolysed *Salmonella minnesota* R7 lipopolysaccharide visible in the electron density. It consists of two heptose saccharides and an anhydro-Kdo residue. Binding by SP-D occurs via the O6 and O7 hydroxyl pairs of HepI. HepIII, which was not visible in the map, is linked to HepII via the O7 group, labelled with the red asterisk. The outer core region of LPS that is not present in this strain is connected via the O3 group of HepII, labelled with the blue asterisk.

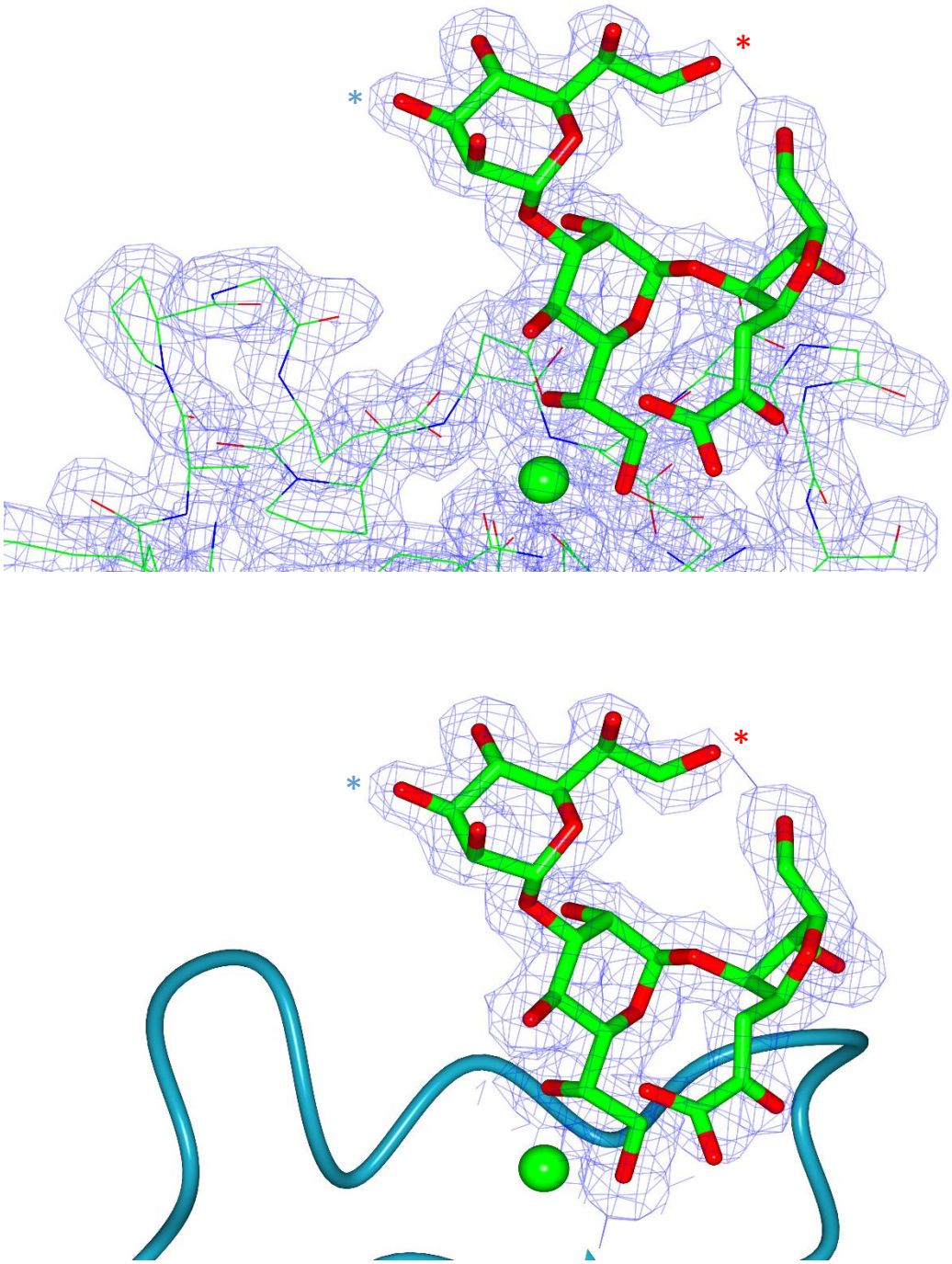


Figure 4-16 Top - 2Fo-Fc map of bound *S. minnesota* R7 polysaccharide in the binding site of the B chain. Bottom - 2Fo-Fc map clipped round the ligand only. The map was contoured at 1σ after restrained refinement of the structure containing ligand and water. HepIII, which was not visible in the map, is linked to HepII via the O7 group, labelled with the red asterisk. The outer core region of LPS that is not present in this strain is connected via the O3 group of HepII, labelled with the blue asterisk. Images generated using CCP4MG.

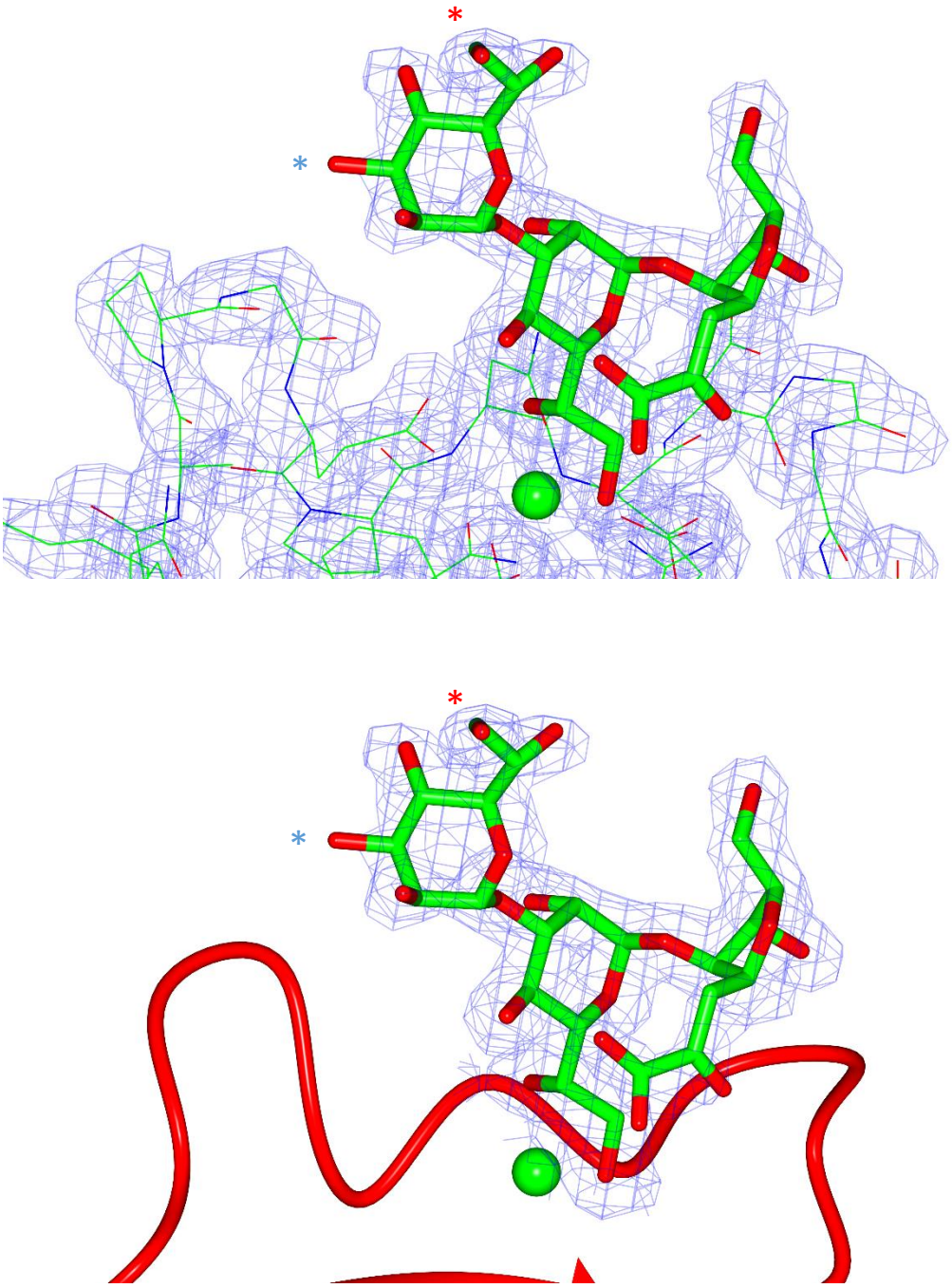


Figure 4-17 Top - 2Fo-Fc map of bound *S. minnesota* R7 polysaccharide in the binding site of the the C chain. Bottom - 2Fo-Fc map clipped round the ligand only. The map was contoured at 1σ after restrained refinement of the structure containing ligand and water. HepIII, which was not visible in the map, is linked to HepII via the O7 group, labelled with the red asterisk. The outer core region of LPS that is not present in this strain is connected via the O3 group of HepII, labelled with the blue asterisk. Images generated using CCP4MG.

The ligand was coordinated mainly via HepI, by sidechains that also coordinate the calcium 1 ion - the OE2 atom of Glu321, ND2 of Asn323, OD1 atoms of Asn323 and Asn341 interact with the O6 hydroxyl group on the glycerol side chain, while the OE2 of Glu329, ND2 of Asn341 and the main-chain carbonyl of Asp342 interact with the O7 group (distances are shown in Table 4-6).

Interactions were also seen between the furanose O6 and tail O1 hydroxyl groups of anKdo - with sidechains of Asp325 and Arg343, respectively. Hydrogen bond distances between the O6 hydroxyl group and OD2 atom of Asp325 were 2.78 Å in the B chain and 2.69 Å in the C chain; Bonds between the O1 tail hydroxyl and NH2 atom of Arg343 were 2.69 Å and 2.45 Å.

Table 4-8 Coordination of ligand by sidechains in the CRD of chains B and C. Distances in Å

Ligand OH group	Protein atom	B chain	C chain
HepI O6	Glu321OE2	2.54	2.48
	Asn323ND2	2.96	2.96
	Asn323OD1	3.02	3.02
	Asn341OD1	3.18	3.19
HepI O7	Glu329OE2	2.59	2.51
	Asn341ND2	2.92	3.02
	Asp342O	2.95	2.88
anKdo O1	Arg343NH2	2.69	2.45
anKdo O6	Asp325OD2	2.78	2.69

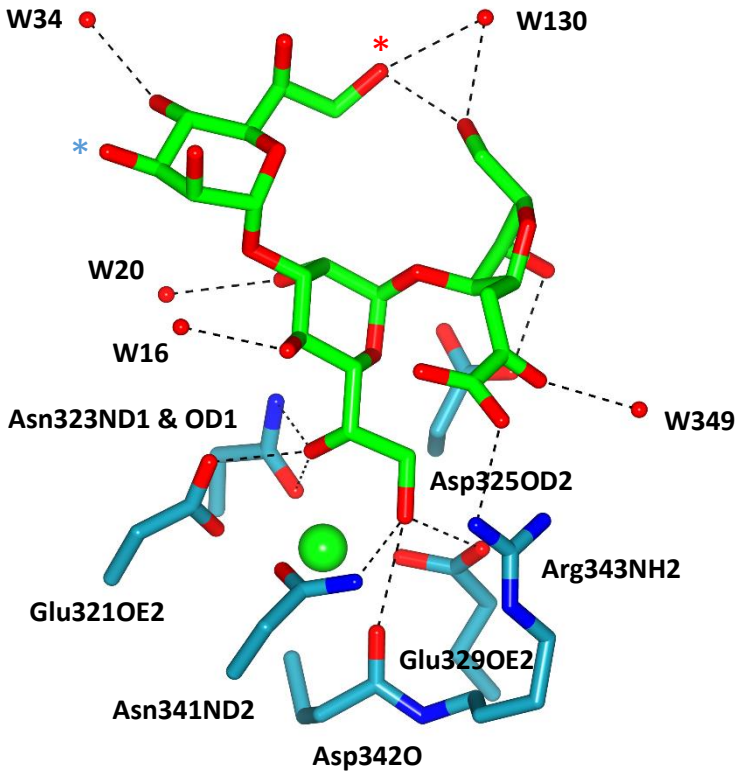


Figure 4-18 Ligand coordinating water molecules and amino acid sidechains in the B chain CRD. The protein is shown in cyan, water molecules as red spheres and the bound *S. minnesota* R7 polysaccharide in green. Labels show the amino acid sidechain atoms and water molecules that interact with the ligand. HepIII, which was not visible in the map, is linked to HepII via the O7 group, labelled with the red asterisk. The outer core region of LPS that is not present in this strain is connected via the O3 group of HepII, labelled with the blue asterisk. Image generated using CCP4MG.

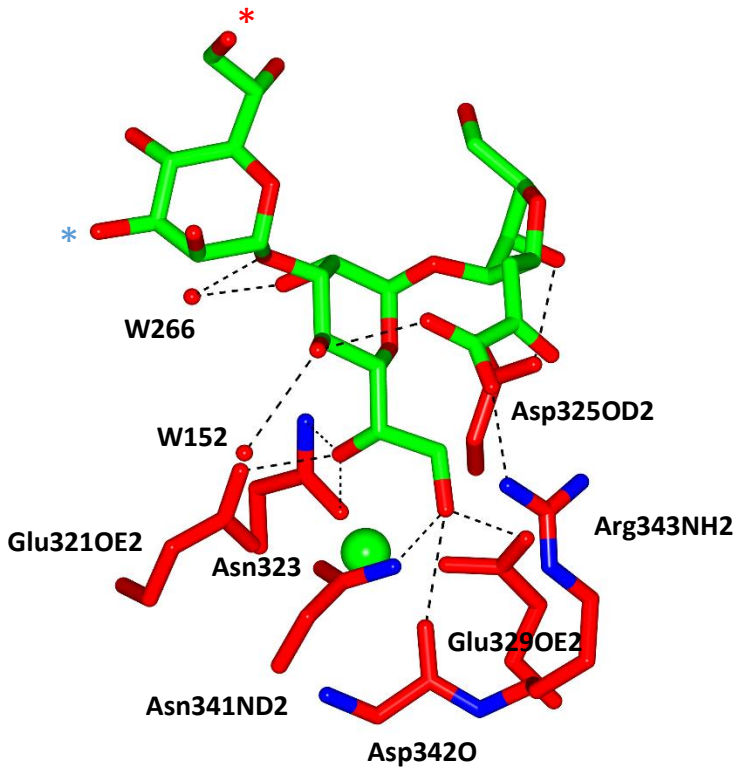


Figure 4-19 Ligand coordinating water molecules and amino acid sidechains in the C chain CRD. The protein is shown in red, water molecules as red spheres and the bound *S. minnesota* R7 polysaccharide in green. Labels show the amino acid sidechain atoms and water molecules that interact with the ligand. HepIII, which was not visible in the map, is linked to HepII via the O7 group, labelled with the red asterisk. The outer core region of LPS that is not present in this strain is connected via the O3 group of HepII, labelled with the blue asterisk. Image generated using CCP4MG.

Notably, binding was exclusively mediated by the side chain of HepI and the O6 and O1 groups of anKdo - and no direct interactions were observed between the protein and the second heptose, HepII, in either chain. The conformation of this residue consequently differs between the B and C chain; there is a 44° rotation round the O1-C1 bond, as well as a slight rotation of the pyranose ring of HepII in the C chain with respect to HepI. There is also a significant 175° rotation of the C6-C7 glycerol side-chain away from the binding pocket (see Fig 4-20).

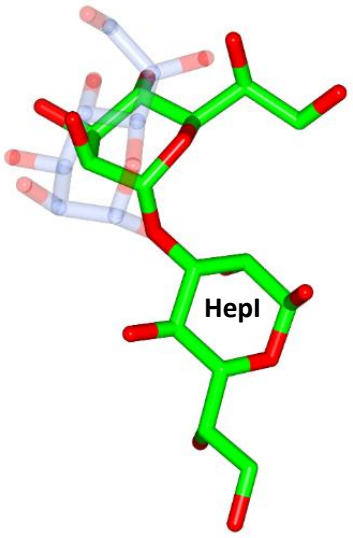


Figure 4-20 Conformational change of HepII relative to HepI in the B chain (solid green backbone) and C chain (semi-transparent ice-blue backbone). HepII in the C chain is rotated round the O1-C1 bond by roughly 44° compared to that in the B chain, as well as a slight rotation of the pyranose ring and a 175° rotation of the glycerol sidechain. Image generated using CCP4MG.

Interestingly, and in addition to the primary protein binding site, sidechains of the symmetry-related trimer were also in close proximity to the ligand. Interactions were weaker than those seen in the CRD and involved two sidechains in the C subunit of the neighbouring trimer interacting with ligand in the B chain, and sidechains in the A chain of the trimer interacting with ligand in the C chain.

Table 4-9 Coordination of ligand by symmetry-related sidechains. Distances in Å

Ligand OH group	Protein atom	B chain	C chain
HepI O2	CSer226OG	3.17	-
HepII O4	CLys229NZ	2.94	-
HepII O4	AGly241O	-	3.10
HepII O6	ASer239OG	-	3.00

In addition to direct interactions between the symmetry-related protein and ligand, a total of six hydrogen-bond bridges are formed between the ligand, water molecules and the neighbouring protein in the crystal form. These bridges are exclusively seen in the B chain, and involve the non-bound HepII and anKdo residues only.

Table 4-10 Hydrogen bond bridges between the ligand and symmetry-related trimer. Distances in Å.

Ligand atom	Water molecule	Protein atom	Distance 1 Ligand-water	Distance 2 Water-protein
HepII O2	W10	CSer226O	2.79	2.88
HepII O2	W140	CGly241O	3.27	2.55
HepII O3	W140	CGly241O	3.24	2.55
HepII O4	W34	BTyr228OH	2.59	2.69
HepII O7	W130	AGln263O	3.02	2.81
anKdo O8	W130	AGln263O	2.89	2.81

The position and orientation of the ligand resulted in interactions occurring between residues that differed between chains. In the B CRD-bound ligand a 2.39 Å hydrogen bond is formed between the HepII O7 hydroxyl group and the O8 hydroxyl group of the non-adjacent anKdo; in the C chain a 2.77 Å hydrogen bond is formed between the O4 hydroxyl group of terminally bound HepI and the O1 tail group of the adjacent anKdo monosaccharide (see Fig 4-21).

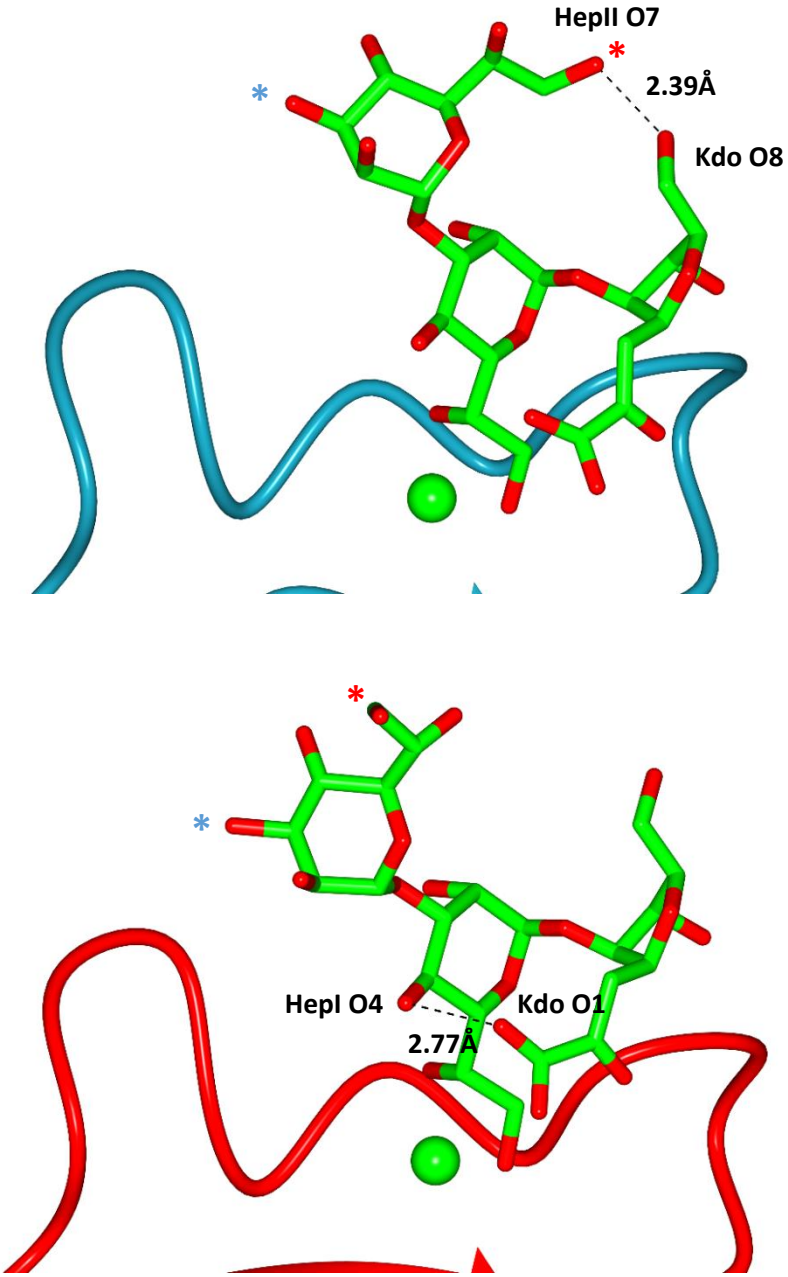


Figure 4-21 Showing interactions that form between residues of the ligand. The top image, in the B chain CRD, shows a 2.39 Å hydrogen bond formed between the O8 hydroxyl group of anKdo and the O7 hydroxyl group of non-adjacent HepII. The bottom image, in the C chain CRD, shows a 2.77 Å hydrogen bond formed between the O1 tail hydroxyl of anKdo with the O4 hydroxyl group on the pyranose ring of neighbouring HepI. HepIII, which was not visible in the map, is linked to HepII via the O7 group, labelled with the red asterisk. The outer core region of LPS that is not present in this strain is connected via the O3 group of HepII, labelled the blue asterisk. Image generated using CCP4MG. Image generated using CCP4MG.

Table 4-11 Refinement statistics for the SP-D - *S. minnesota* R7 complex (crystal CCS16A11)

Refinement	
Resolution range (Å)	55.65 – 1.77
R-factor (%)[a]	18.71
R-free* (%)[b]	22.05
Protein atoms	3466
chain A	1154
chain B	1154
chain C	1149
Other atoms	
Calcium ions	9
Ligand	82 (41 in B, 41 in C)
Waters	386
Average B values (Å^2)	
Protein main chain	22.8
Calcium ions	17.1
Ligand	25.5
Water	33.3
Ramachandran plot values (%)	
Most favoured	97.31
Additional allowed	2.47
Disallowed	0.22

[a] $R_{conv} = \Sigma_h ||F_{oh}| - |F_{ch}|| / \Sigma_h |F_{oh}|$, where F_{oh} and F_{ch} are the observed and calculated structure factor amplitudes, respectively, for the reflection h.

[b] R_{free} is equivalent to R-factor for a 4.87% subset of reflections not used in the refinement.

Chapter 5: Discussion

5.1 Crystal growth

The late addition of calcium to pre-existing sitting drops of protein preparation 3 - which had undergone affinity purification and size-exclusion chromatography as well as additional endotoxin removal - was the only condition to produce crystal growth. Whether successful growth was a consequence of purer protein, or due to the late addition of calcium is unknown.

Gram-negative bacteria are widely used to produce recombinant protein and endotoxins (LPS) are major contaminants of recombinant protein production. Although found on the bacterial cell wall, LPS is continuously liberated and its removal is often crucial when producing a sample of sufficient purity. In addition to isolation and purification of the macromolecule itself, preparation 3 was the only sample to be treated with endotoxin. As previously stated, a real crystal is rarely a single crystal, but rather a mosaic of near perfectly aligned domains. Misalignments of no more than a few tenths of a degree are useful for diffraction, but in some cases contamination may be severe enough to hinder crystal growth entirely.

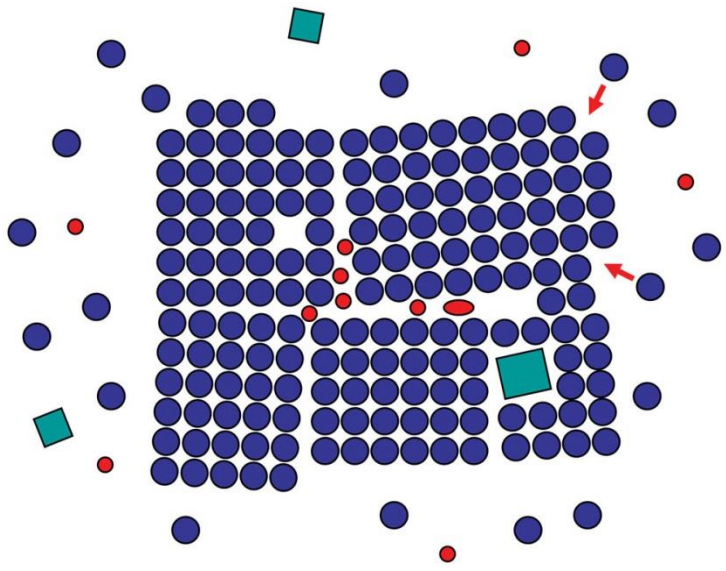


Figure 2-4 (replicated) Schematic drawing of a highly mosaic crystal growing in solution. Protein molecule (blue spheres) are interrupted by small impurities (red) and debris (green). New molecules preferentially attach to ledges or steps, shown with red arrows. Contamination occurring within the growing crystal causes defects and misalignment, in this case of around 6°. Image taken from Rupp, 2010.

Endotoxin treatment of the protein sample was, in this case, necessary for formation of diffraction quality crystals. This same preparation which had been dialysed in to calcium-containing buffer also failed to produce crystals.

Slowing the rate of crystal growth is a common method of obtaining larger, and superior quality crystals. Typical approaches include addition of reagents that solubilise the sample - such as detergents and salts; coating the reservoir solution with paraffin or silicon oil to reduce the rate of vapour diffusion between it and the drop; or by diluting the drop itself, which supresses excessive nucleation (Blow *et al.,* 1994). It has also been noted that addition of cofactors or ligands can modify the crystallisation behaviour of proteins - and this may be the case when adding calcium late; certain non-protein chemical compounds that are bound to macromolecules are often required for biological activity, and crystallisation may require their presence in solution for crystal nuclei to form (Borisenko & Kolesnikov, 2012).

Whilst the metastable zone is considered to be an optimal zone for crystal growth, it may be possible to sustain but not initiate growth. Before addition, no calcium was present in the drop and as SP-D is a calcium dependent C-type lectin - It can be inferred that the presence of calcium is not only necessary for carbohydrate binding, but also for protein stabilisation. Amino acids Asp325 and Arg343 are directly implicated in forming crystal contacts, and are located within close proximity to the binding pocket and the carbohydrate-binding Ca1 ion (see Fig 5-1). It may be plausible that a lack of calcium at this site causes increased flexibility of the surrounding protein residues, and decreased ability to form the crystal contacts necessary for the protein to aggregate and a crystal to grow.

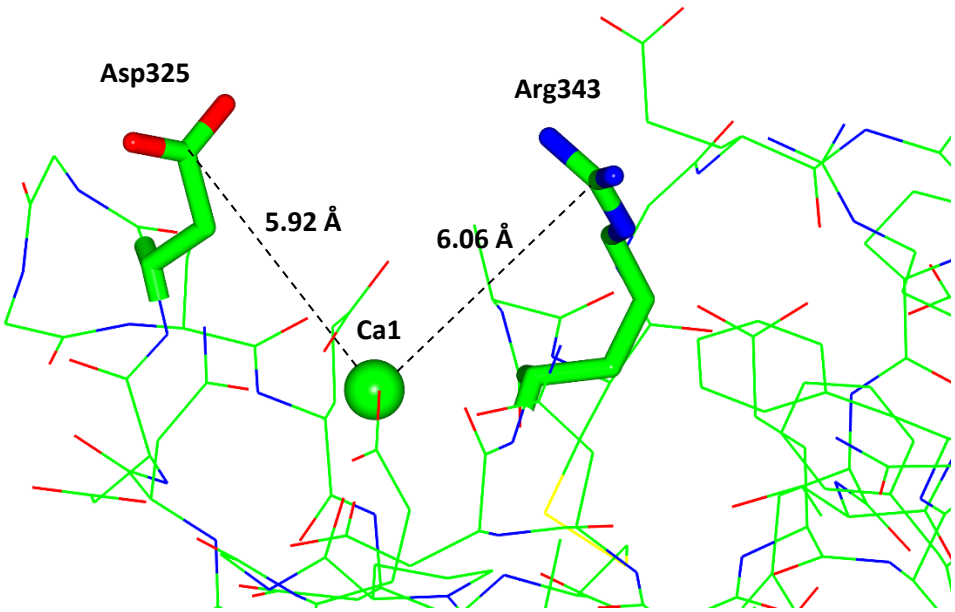


Figure 5-1 Showing the proximity of Asp325 and Arg343 to the ligand-binding Ca1 ion in the CRD of the B chain. Distances to the primary carbohydrate binding calcium ion, Ca1, are also shown. Image generated using CCP4MG.

Unlike the trials presented here, previously solved structures of SP-D were crystallised by Shrive *et al.,* 2003; Shrive *et al*., 2009 in sitting drops consisting of equal volumes of protein solution and precipitant buffer, with the protein in a buffer containing 10mM $CaCl_2$. The protein stock provided was produced in the same manner as that used here, but at a different time and date and by a different research group - which may account for differences seen during crystallisation attempts. Sitting drops used to grow SP-D crystals also consisted of larger volumes of protein and precipitant buffer - 5µl - compared to 1µl here. Different volumes of protein and precipitant may be tried as part of the optimisation process highlighting the temperamental nature of crystal growth. For instance, using more protein may result in a net increase in concentration of the protein in the drop after equilibrium is reached, which can be desirable for dilute protein samples, or to slow crystal nucleation or growth (Dessau & Modris, 2011). It is interesting that only this batch of protein yielded crystals so unconventionally with late addition of calcium a serendipitous

after-thought - when crystallisation trials of other protein batches in calcium buffer before and after this as described by Shrive *et al.,* 2003; 2009 produced crystals of equal quality.

Once the late addition procedure was discovered, and tests proved it to be reproducible in extensive screening round the successful conditions, certain wells repeatedly yielded crystal growth. Conditions that consistently produced diffraction quality crystals included 16% weight/volume PEG 10,000, and 0.1M Tris at pH7, as well as 16% weight/volume PEG 4,000, and 0.1M Tris at pH8.

5.1.1 Cocrystallisation trials

Cocrystallisation is a common method of obtaining crystals of a protein-ligand complex, and is often the technique of choice when the compounds are insoluble or aggregate easily (Hassell *et al.*, 2006). There are however, many factors that may influence successful cocrystallisation including temperature, whereby incubation may facilitate complex formation; and protein/ligand concentration - as sometimes it may be possible to concentrate the protein then add the ligand to form a complex, but insoluble ligands may cause the protein to precipitate at higher concentrations. None of cocrystallisation trials yielded crystals. In each well of trays CCS7 and CCS14 and wells C1 and D1 of CCS16, thick brown precipitate formed and no crystal growth was seen.

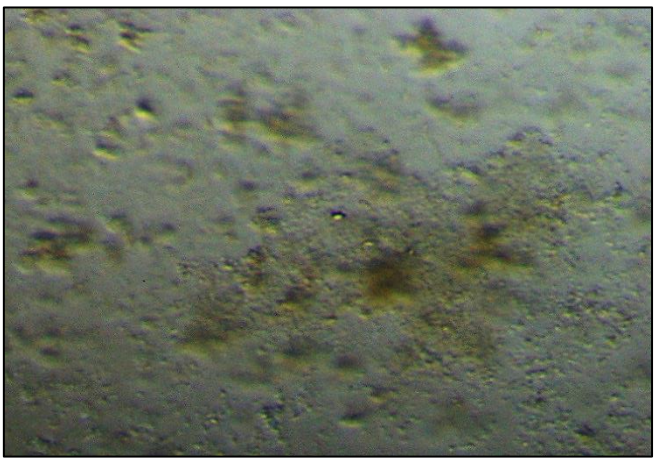


Figure 4-1 (replicated) Well CCS7D2 showing typical formation of cloudy precipitate during cocrystallisation.

All cocrystallisation samples were allowed to incubate and were laid down at room temperature. Heat-treating the protein-ligand is a technique that may improve the quality of the protein by reducing macromolecules that may not be properly folded, giving a more homogenous sample. Typically, 1:3 to 1:10 protein to ligand ratios are used to achieve stable complexes that yield well-diffracting crystals (hamptonresearch.com, 2015). All trials (excluding CCS7 which contained diluted protein that was concentrated along with intact *H. influenzae* Eagan 4A) consisted of a protein to ligand ratio of 1:4 to 1:5, with calcium present in the solution. This may have been too high a ligand concentration, causing the protein to precipitate out.

Tray CCS7 contained protein that had been incubated with intact LPS, including the hydrophobic and insoluble lipid A moiety. LPS molecules have limited or no water solubility, and successful complexation may have been improved by dissolving the ligand in a solubilising agent such as DMSO (dimethyl sulphoxide) or a small amount of low molecular weight polyethylene glycol.

5.2 Unsuccessful ligand soaks

Unlike the aforementioned cocrystallisation trials which saw no crystal growth at all, other trays yielded growth but showed no ligand binding after soaks and cryoprotection. Once data was collected and maps were calculated, no electron density was present in the binding site for

crystals soaked with intact *H. influenzae* Eagan 4A LPS, hydrolysed smooth LPS from *E. coli* B4 and B6, and rough LPS from *E. coli* F583. As previously mentioned this ligand contains a large, hydrophobic lipid A moiety which is not water soluble (see Fig 5-3) - and either was too insoluble to dissolve in to the protein solution or was too large to enter the crystal. Packing of protein molecules in the crystal forms channels through which a ligand must pass to enter the ligand binding pocket (see Fig 5-2).

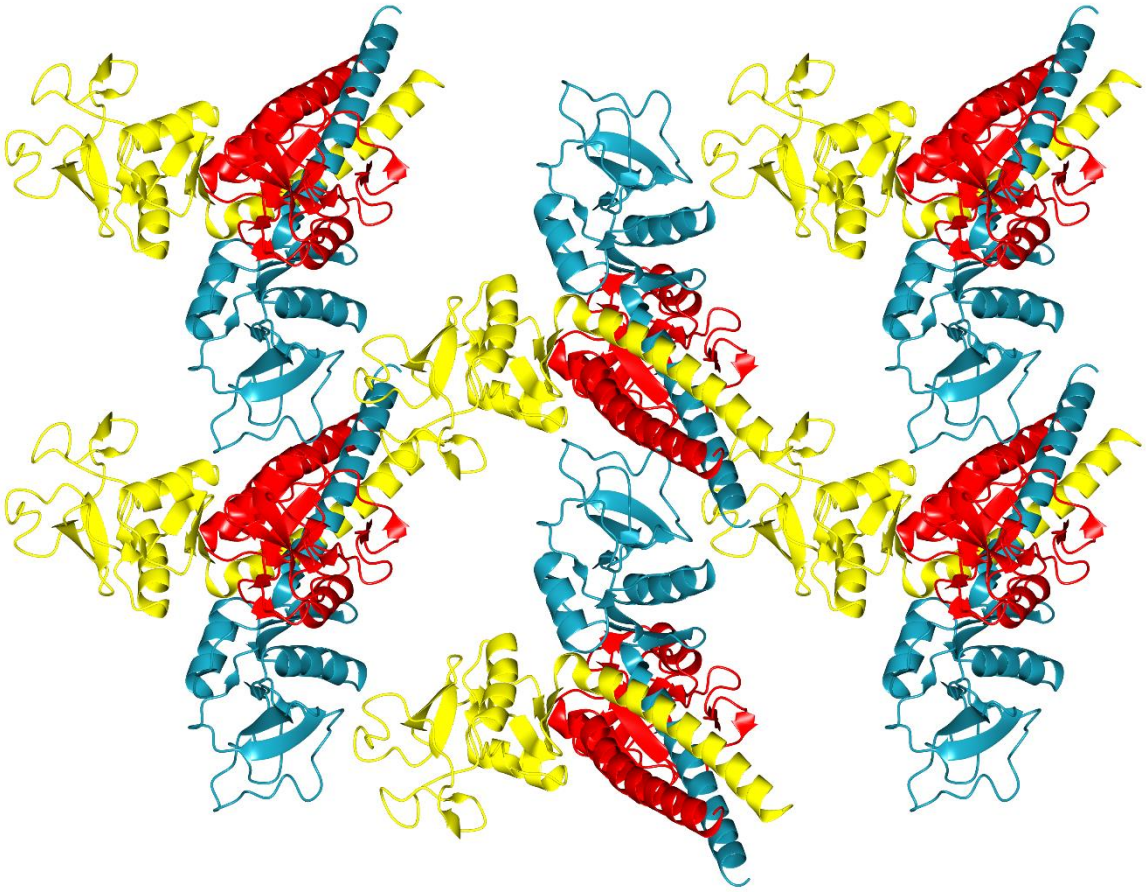


Figure 5-2 Showing protein molecule packing within a crystal of SP-D, viewed along the BC face. Due to the formation of crystal contacts in close proximity to the binding pocket, access to the CRD is only possible if a ligand passes through a series of channels formed by the symmetry-related protein. Chain A is shown in yellow, chain B in cyan and chain C in red. No ligand or calcium ions are shown. Image generated using CCP4MG.

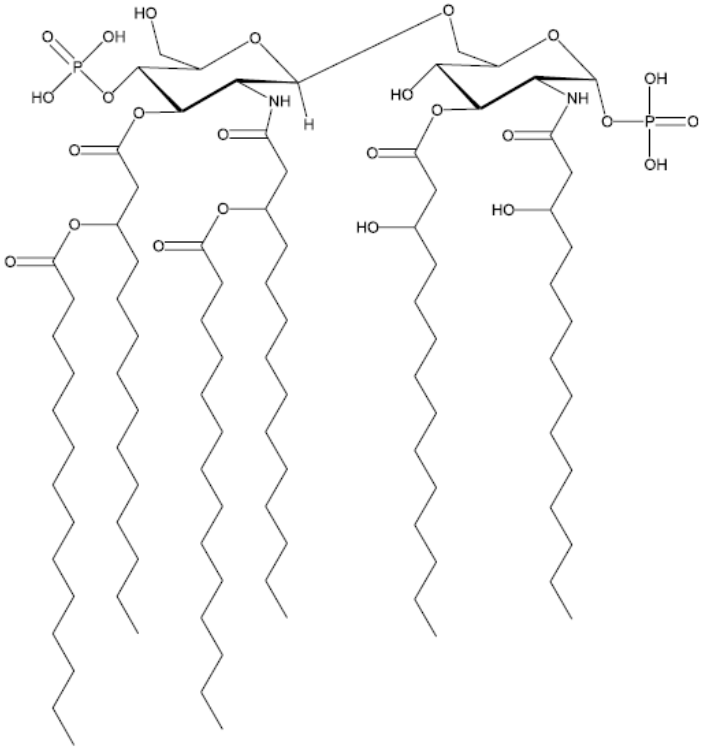


Figure 5-3 General structure of the diphosphorylated lipid A moiety from *H. influenzae* LPS. It consists of a double-glucosamine backbone, each with two amide-linked 3-hydroxytetradecanoic (myristic) acids and a single phosphate group. A single myristic acid residue measures approximately 17.4 Å in length as measured in Coot. Image taken from Mikhail *et al.,* 2005.

Just as glycosylated proteins are notoriously difficult to crystallise, it can be inferred that longer ligands would be too. Typically, crystallisation requires removing floppy ends, disordered loops or intrinsically unstructured (natively unfolded) regions – to minimise flexibility and disorder. Smooth strains of lipopolysaccharides contain repeating oligosaccharide O-antigens, with an average molecular weight six-times that of rough LPS (Backhed *et al.,* 2001) - and no smooth LPS tested during soaks was visible in any structure.

Smooth *E. coli* B4 and B6 LPS strains were initially chosen for soaks due to SP-D's preferential binding to heptose residues, and that the outer core and O-antigen would effectively be ignored. This was not the case however, and no evidence of binding was seen.

E. coli O111:B4 LPS strains are believed to contain a full inner and outer core and between 18 and 27 repeating O-antigen units (see Fig 5-4) with a molecular weight of roughly 7,000Da (Morrison & Leive, 1975) and SP-D has been shown to exhibit significant binding to the strain in solid-state binding assays (Yamazoe *et al,* 2008). Even without the large, hydrophobic lipid A moiety, the ligand may have been too large to successfully penetrate the crystal channels and access the binding pocket.

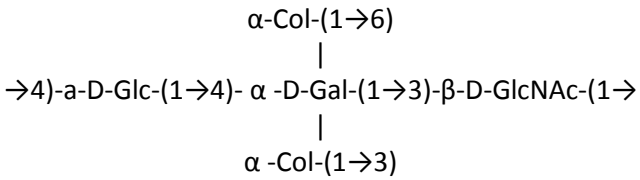


Figure 5-4 Structure of the O-antigen repeat unit of *E. coli* B4 LPS. Glc - glucose; Gal - galactose; Col - colitose; GlcNAc - N-acetyl glucosamine. Colitose is a carbohydrate product of mannose and is found in the O-antigen of certain Gram-negative bacterial LPS such as *E. coli*, *Yersina*, *Salmonella* and *Vibrio* species (Samuel & Reeves, 2003). Structure taken from Stenutz *et al*., 2006.

E. coli O26:B6 exhibits short-chain length behaviour on SDS-PAGE and shows low molecular weight bands and low critical micelle concentration which is considered to be directly proportional to polysaccharide chain length (Mangoni *et al.,* 2008; Aurell & Winstrom, 1998). Peterson *et al*., 1985, separated various smooth *E. coli* strains by gel filtration in to short- and long-chain fractions containing an average of 1 and 18 O-antigen repeat units respectively. By demonstrating short-chain-like behaviour, the B6 strain LPS is likely to contain only one repeat unit, shown below (see Fig 5-5). SP-D has shown significant binding to the strain in solid-state binding assays (Leth-Larsen *et al.,* 2005; Yamazoe *et al.,* 2008), but the whole ligand may be too large to successfully penetrate the crystal channels and access the binding pocket.

→2)-β-D-Man-(1 →3)- β -D-Gal-(1 →3)- α -D-GlcNAc-(1 →3)- β -D-GlcNAc-(1→
|
α -L-Fuc-(1→2)

Figure 5-5 Structure of the O-antigen repeat unit of *E. coli* B6 LPS. Man - mannose; Gal - galactose; GlcNAc - N-acetyl glucosamine; Fuc - fucose. Structure taken from Stenutz *et al*., 2006.

No density was seen in the ligand binding site for F583-soaked crystals. This ligand, known to exist as a small Rd_2 mutant, has not been well characterised however, and is believed to comprise only of Kdo and a single heptose residue - HepI. Crystals deteriorated rapidly during tests and were frozen very quickly after exchange, suggesting that the ligand may be disrupting interactions between protein molecules in the crystal. This may be plausible if the ligand is small enough to pass through the solvent channels in the crystal, but contains additional groups such as negatively charged phosphates that may alter the crystal environment. SP-D was shown to bind Rd_1 mutant *S. minnesota* R7 which contains two additional heptose residues that would not interact with the protein, compared to that of F583. Given the structurally conserved inner core of *Salmonella* and *E. coli* LPS, it can be presumed that binding by SP-D would occur in a similar manner - that is, by heptose HepI. Also only 10mM F583 ligand was present in the cryobuffer - significantly lower than other tests and less than half that of successful *S. minnesota* R7 soaks.

5.3 Overall structure of R7-bound SP-D

The overall structure of the recombinant head and neck fragment of human SP-D reveals a trimeric structure of three C-terminal globular head domains, linked by an extended α-helical coiled coil neck region, with a short N-terminal collagen domain that is not visible in the electron density. The structure contains residues 205-355 in chains A and B, and 206-355 in chain C.

The head and neck domains of the protein were very well defined, as were the locations of known calcium ions. Calcium presence was clearly visible in the electron density for sites 1, 2

and 3 in the CRD, with a marked absence at the fourth location in the central pore (see figure 4-13, replicated below).

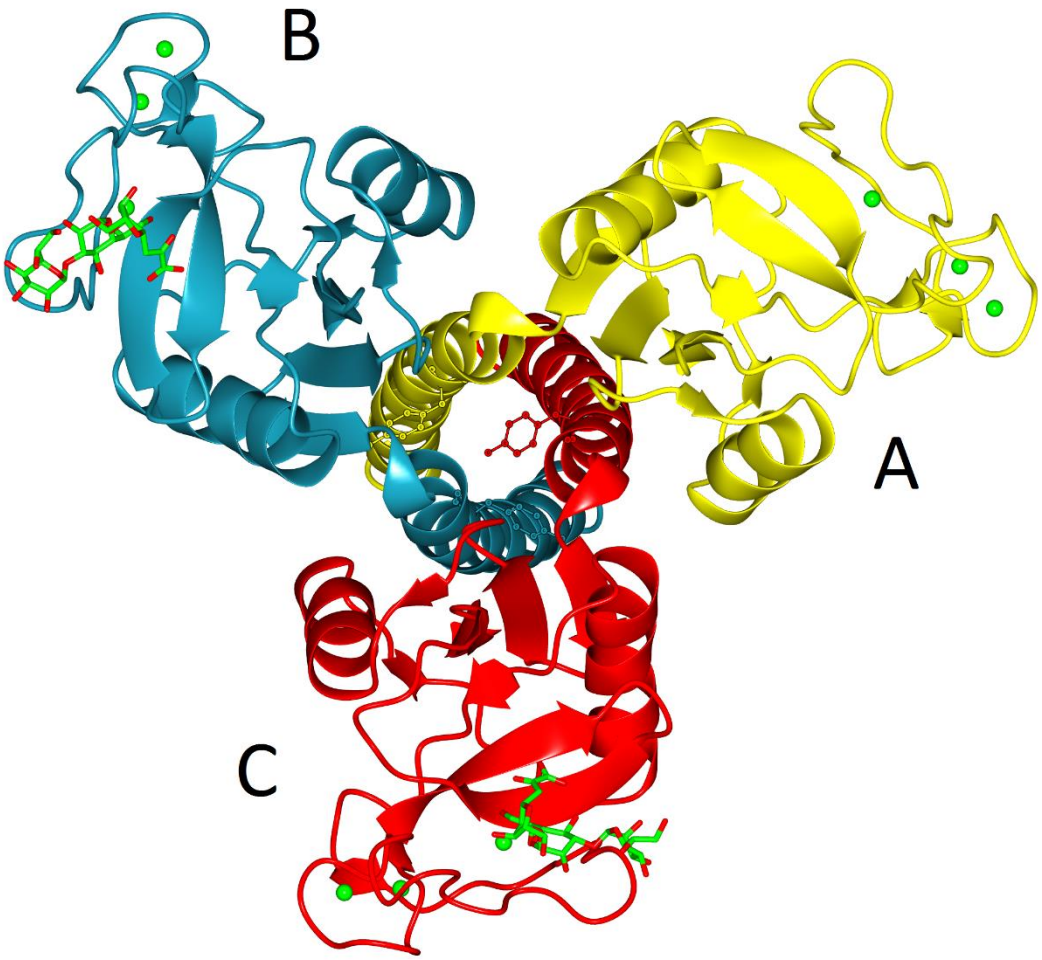


Figure 4-13 (replicated) Crystal structure of ligand-bound SP-D in ribbon form, showing three calcium ions per CRD in green, and the asymmetric tyrosine positioned into the central pore. Ligand is seen in the B and C chains. Image generated using CCP4MG.

The absence of Ca4 at this site has previously been demonstrated in the maltose-bound structure elucidated by Shrive *et al.,* 2003, and is consistent with lower concentrations of calcium ions in the cryoprotectant. The final calcium concentration was ~0.17mM due to an absence of calcium in the cryoprotective agent. Other factors that may contribute to this absence may include the presence of high concentrations of R7 LPS ligand (22.2mM) and the varying pH and PEG molecular

weight in the crystallisation conditions and cryobuffer - which may alter the electrochemical environment within the crystal that is otherwise conducive to calcium occupancy at certain sites in the protein. These results (coupled with structures that have this ion present) suggest that a minimum calcium concentration greater than that included in the protein solution or cryoprotectant is required for occupancy of the Ca4 site (Shrive *et al*., 2003).

Also concurrent with the studies by Shrive *et al*., is the resulting asymmetry of residues Glu232 and Tyr228 in the Ca4-depeletd structure. The position of the C chain tyrosine sidechain into the centre of the coiled-coil allows it to interact with one of the three glutamate residues at position 232 (at a distance of 2.42 Å), that otherwise interact with glutamates at this position in other subunits and with Lys246 in the same chain. The intrachain interactions of Glu232 and Lys246 between the neck and head region of chains A and C are relatively strong (2.99 Å and 2.91 Å, respectively) compared to that in the B chain (5.59 Å), and suggest a significant charge transfer to the two lysine residues. It is unlikely that this fourth calcium ion is present in other collectins other than bovine SP-D, since no acidic residues equivalent to Glu232 are present; nonetheless, a basic residue at position 246 is a common feature of SP-D proteins across species. The asymmetry of Tyr228 and resulting asymmetry of Glu232 does not disturb the geometry of the neck-CRD interface.

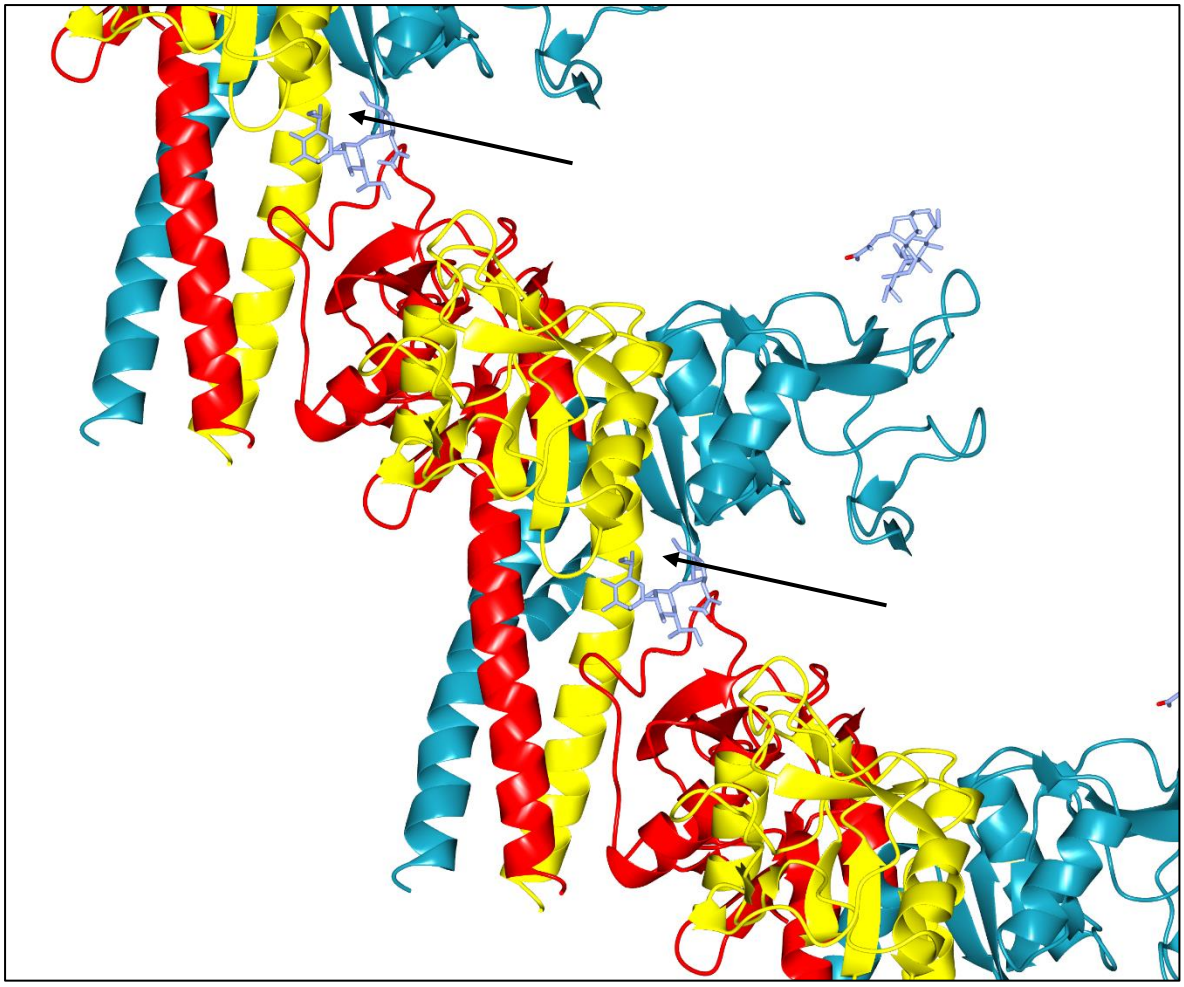


Figure 5-6 Showing the close proximity of the neck region to the ligand binding site of a symmetry-related trimer. Arrows indicate the contact region. Chain A is shown in yellow, chain B in cyan and chain C in red. Ligand is shown here in the C chain only, in ice-blue. Image generated using CCP4MG.

An interesting feature that may support the suggestion that asymmetry in the neck occurs as a result of crystal packing is the close proximity of the ligand and the asymmetric tyrosine at position 228 in the same chain in a symmetry-related trimer (see Fig 5-7) that may help to stabilise the ligand in the crystal and define it in the electron density.

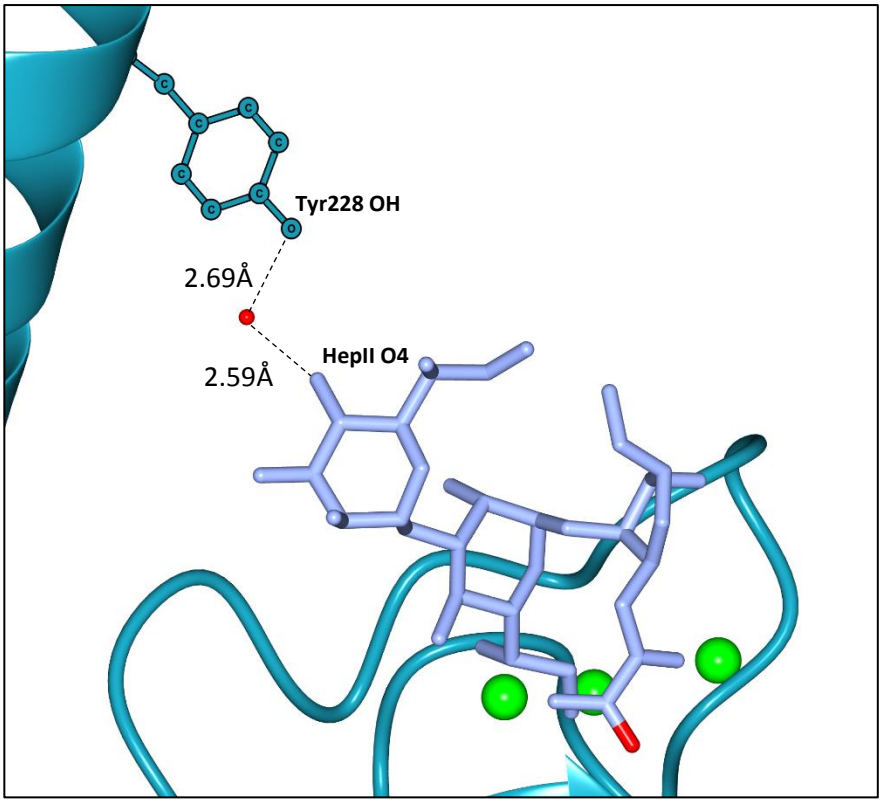


Figure 5-7 Showing the close proximity interaction between the hydroxyl group of tyrosine 228 in the B chain and the O4 group of HepII from a symmetry-related trimer. Protein chain B of both molecules is shown in cyan ribbon-form, with the protruding tyrosine in circles. The ligand in the CRD is shown in ice-blue, with three calcium ions in green. There is a hydrogen bond bridge formed between the ligand, a water molecule (red sphere) and symmetry-related trimer, due to the protein molecule packing within the crystal. Image generated using CCP4MG.

5.3.1 The ligand binding site

As previously mentioned, SP-D demonstrates a significant ability to recognise natural ligands such as bacteria and viruses, as well as cell surface receptors (Gardai *et al.,* 2003) - and the structure presented in this thesis provides insight in to the mechanism of bacterial lipopolysaccharide binding at the primary binding site in the CRD.

The two heptose residues visible in the electron density conform to the structure previously solved by Wang *et al.,* 2008 (see Fig 5-8 and Table 5-1), whereby the inner-most heptose (HepI) is bound in a calcium-dependant manner and interacts directly with the protein. HepI is bound in a specific manner via the glycerol tail O6 and O7 hydroxyl groups, and is coordinated by residues of

the CRD that also coordinate the Ca1 ion. There was no ligand-protein interaction between the second heptose saccharide (HepII) and the protein - implicating the inner-most heptose as the primary site of the carbohydrate interaction.

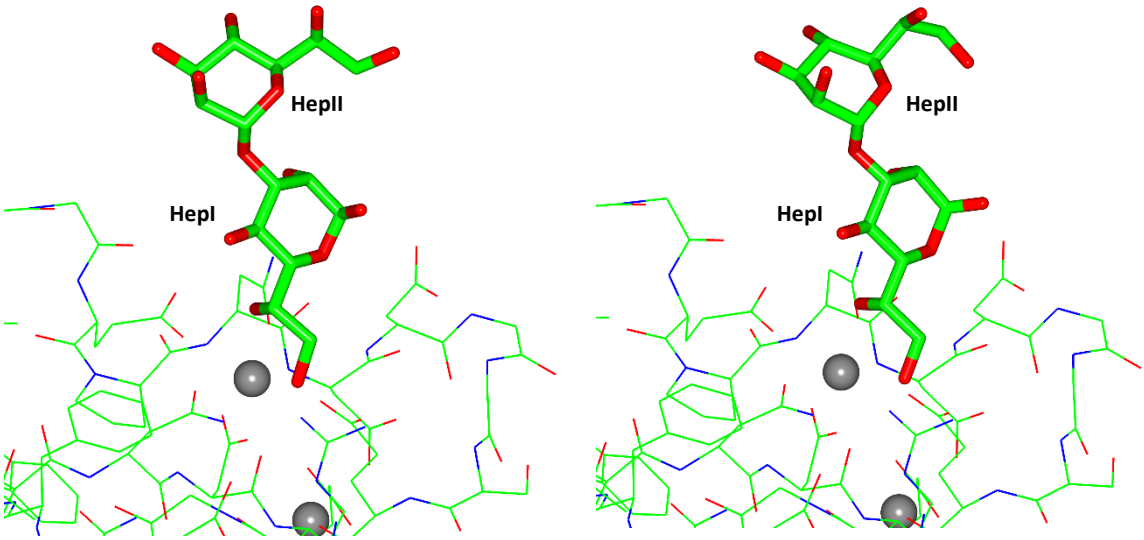


Figure 5-8 Comparison of binding and ligand orientation of heptose disaccharide in the B chain between the R7 data presented in this thesis (left, Kdo residue not shown), and the structure previously reported by Wang *et al.*, 2008 (right, PDB accession code 2RIC). Binding is essentially identical with regards to position and coordination of HepI by the calcium ion Ca1, with a slight rotation round the glycosidic bond. Calcium ions are shown as grey spheres. Image generated using CCP4MG.

Table 5-1 Comparison of ligand coordination in the B chain between the R7 data presented in this thesis and the structure previously reported by Wang *et al.*, 2008 (PDB accession code 2RIC). Distances in Å

Ligand OH group	Protein atom	Distance (R7 data)	Distance (2RIC)
HepI O6	Glu321OE2	2.54	2.61
	Asn323ND2	2.96	3.07
	Asn323OD1	3.02	3.33
	Asn341OD1	3.18	3.28
HepI O7	Glu329OE2	2.59	2.62
	Asn341ND2	2.92	3.08
	Asp342O	2.95	3.26

This binding mechanism leaves hydroxyl groups available for participation in glycosidic linkages to Kdo residues and longer chain saccharides, when placed in the context of full-length, physiological LPS (see Fig 5-9 and Fig 1-8, replicated below).

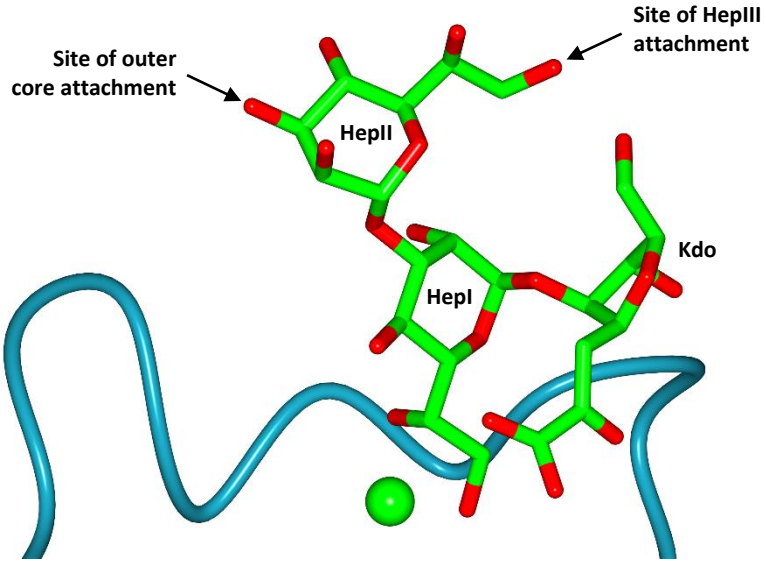


Figure 5-9 Model of the ligand in the B chain CRD, showing the sites of attachment for the LPS outer core not present in this strain, and that of HepIII which is not visible in the electron density map. Image generated using CCP4MG.

The figure below represents the remaining portions of LPS not present in the R7 strain:

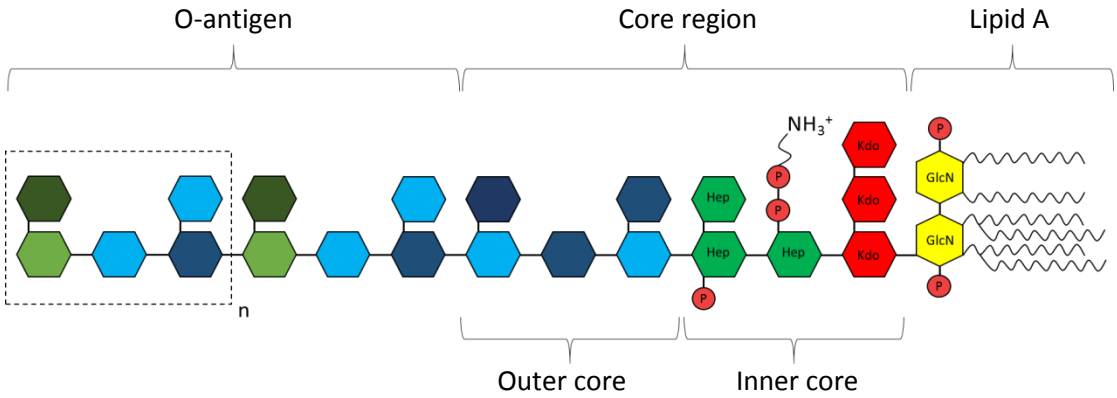


Figure 1-8 (replicated) The general structure of LPS showing hydrophilic region (lipid A), core region made up of an inner and outer portion, and the O-antigen. The terminal structure (hashed box) is the repeating unit of the O-antigen. GlcN - glucosamine; Kdo - 3-deoxy-D-manno-octulosonic acid; Hep - D-glycero-D-manno-heptose. Image adapted from Alexander & Rietschel, 2001.

Considering the restricted size of the binding pocket and close proximity of neighbouring molecules in the crystal (see Fig 5-2), a longer LPS mutant would be too large to access the CRD in the crystal form.

5.3.2 Non-binding pocket interactions

In addition to direct binding of the ligand in the CRD, other interactions occur between symmetry-related trimers in the crystal and within the ligand itself which contribute to contribute to the electron density being better defined in the B chain compared to the C chain (see Fig 4-14 replicated below).

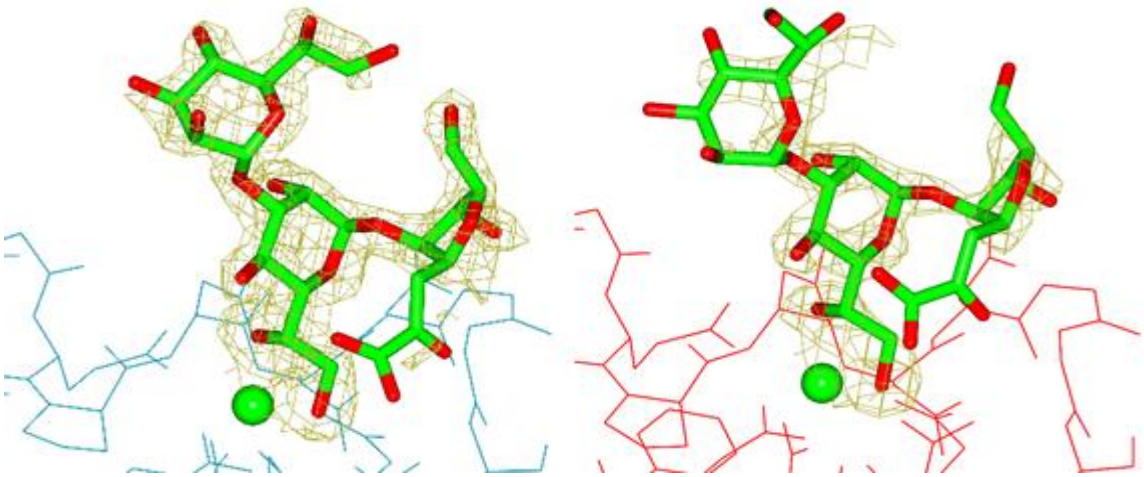


Figure 4-14 (replicated) Fo-Fc difference map of bound *S. minnesota* R7 polysaccharide in chains B (left, blue) and C (right, red), calculated prior to adding ligand or water molecules into the model. The maps were contoured at 2.5σ. Image generated using CCP4MG

5.3.2.1 Symmetry-related protein interactions

Along with the primary site of ligand binding, symmetry-related neighbouring protein trimers also contribute to LPS coordination in the crystal form. As previously stated, molecules packed in the crystal impose constraints on binding of ligands in subunit A, and as such no ligand was visible in the electron density in this chain. Unlike the maltose-bound structure, where the full ligand could only be seen in chain A due to a stabilising water-bridge between the second

saccharide unit and the symmetry-related protein, the longer R7 ligand is hindered from entering the more constricted binding pocket. Sidechains of the symmetry-related trimer are also in close enough proximity to the ligand in chains B and C to interact: Ser226 and Lys229 of the C chain formed weak bonds with heptose residues in the B chain, while Gly241 and Ser239 form weak bonds with heptose in the C chain.

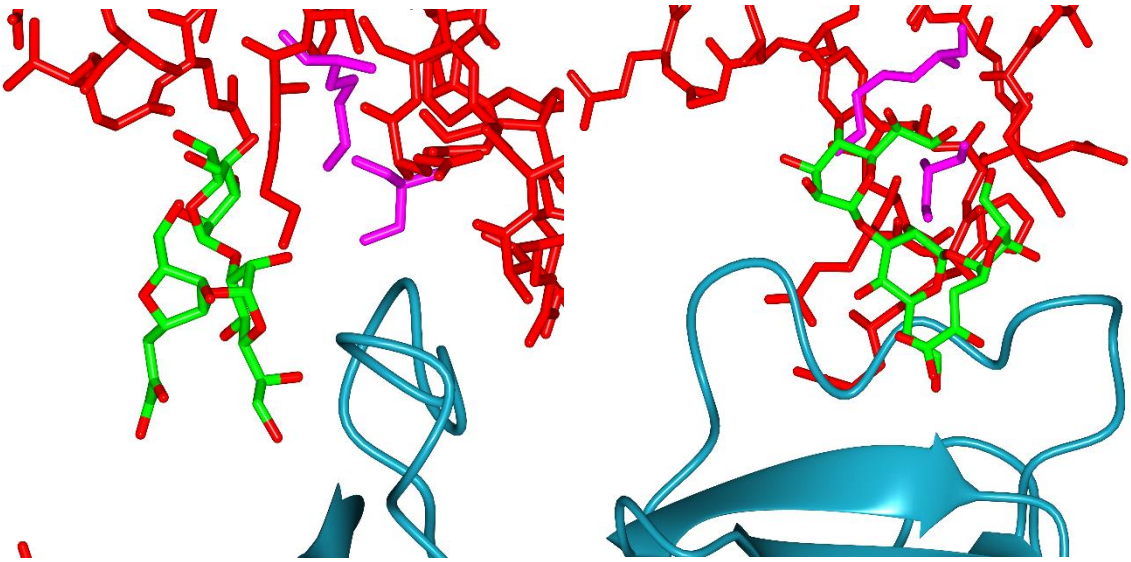


Figure 5-10 Showing interactions that form between residues of the symmetry-related trimer and the ligand in the B chain. The ligand is shown in green, the binding pocket in cyan, and the neighbouring trimer is shown in red. Two residues of the neighbour which form direct bonds with the polysaccharide, Ser226 and Lys229, are magenta. Image generated using CCP4MG.

As mentioned previously, there is also an interaction between the hydroxyl group of tyrosine 228 in the B chain and the O4 group of HepII from a symmetry-related trimer (see Fig 5-7). The electron density round HepII is very well defined in the B chain compared to that in C, but the symmetry-related contacts alone may not explain the reason for this - the cause becomes apparent when examining the hydrogen-bond bridges that form between the ligand and the symmetry-related trimer that are only seen in the B chain (see table 4-10).

Table 4-10 (replicated) Hydrogen bond bridges between the ligand and symmetry-related trimer. Distances in Å.

Ligand atom	Water molecule	Protein atom	Distance 1 Ligand-water	Distance 2 Water-protein
HepII O2	W10	CSer226O	2.79	2.88
HepII O2	W140	CGly241O	3.27	2.55
HepII O3	W140	CGly241O	3.24	2.55
HepII O4	W34	BTyr228OH	2.59	2.69
HepII O7	W130	AGln263O	3.02	2.81
anKdo O8	W130	AGln263O	2.89	2.81

Again, although these interactions arise as a direct consequence of protein packing within the crystal, it does demonstrate SP-D's inherent flexibility and ability to form additional interactions that may occur during agglutination of microbes. It is a well-documented phenomenon that SP-D aggregates microbes to create a clumped mass that inhibits colonisation and enhances clearance by alveolar macrophages. Independent, trimeric subunits display only a weak affinity for saccharides, suggesting that higher-order multimers are required for strong binding and to allow multivalent binding to ligands, or to simultaneously bind both a pathogenic surface structure and receptors to initiate an immune response.

5.3.2.2 Ligand-ligand interaction

In addition to protein-ligand interactions, the R7 ligand also forms intramolecular bonds. It has previously been demonstrated that the extent of heptose branching in lipopolysaccharides of *Haemophilus influenzae* directly influences binding of collectins (Orgeig *et al.,* 2010) - and that higher-order arborisation effectively shields the inner core heptose residues from being bound. It was suggested that bacteria may have adapted by developing surface structures which hinder, or in some cases fully resist, collectin binding - and successfully subvert the innate immune response, but the manner by which they do so was not determined (Orgeig *et al.,* 2010). The interactions between saccharide units of the LPS demonstrated in this thesis are relevant in that they illustrate the ability of longer ligands to fold back, and potentially sterically hinder heptose accessibility. The fact that two distinct examples of these interactions can be seen in the structure, further supports the suggestion.

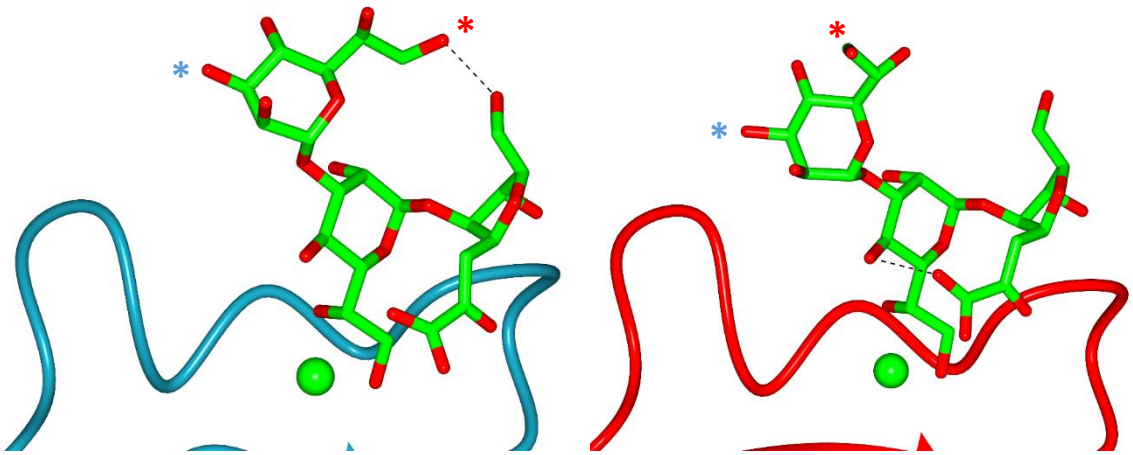


Figure 4-21 (replicated and edited) Showing interactions that form between residues of the ligand. The left-hand image, the B chain CRD, shows a 2.39 Å hydrogen bond formed between the O8 hydroxyl group of anKdo and the O7 hydroxyl group of non-adjacent HepII. The right-hand image, the C chain CRD, shows a 2.77 Å hydrogen bond formed between the O1 tail hydroxyl of anKdo with the O4 hydroxyl group on the pyranose ring of neighbouring HepI. HepIII, which was not visible in the map, is linked to HepII via the O7 group, labelled with the red asterisk. The outer core region of LPS that is not present in this strain is connected via the O3 group of HepII, labelled with the blue asterisk. Intramolecular interactions are shown as hashed lines. Image generated using CCP4MG. Image generated using CCP4MG.

5.3.3 Non-terminal binding

Monosaccharide binding by SP-D has been studied extensively, but the precise mechanism by which longer, physiological polysaccharides are recognised is poorly understood – particularly with regards to binding of non-terminal residues. Studies by Allen *et al*., 2001 used automatic docking and inhibition analysis to demonstrate binding of non-terminal glucosyl residues, but a crystal structure demonstrating such feat is yet to be published. The data presented in this thesis provide an accurate and novel binding mechanism and offer insight in to SP-D recognition of a range of microbial surface carbohydrates – by direct binding to an *internal* carbohydrate residue.

Salmonella Rd1 mutant lipopolysaccharides contain two Kdo units and three inner core heptose residues (Yoshii *et al*., 2013); but only the first two heptoses are visible in the map presented in this thesis, with a marked absence of a third. As HepII shows no direct interaction with the protein, and the orientation of the O7 hydroxyl group of HepII (that HepIII is linked to) is angled away from the CRD, it is likely that a further heptose would also show no direct binding (see Fig 5-10). The O7’ group of the glycerol tail is not coordinated by any direct interaction with the protein, allowing conformational flexibility round the HepII-HepIII bond. This flexibility renders HepIII too mobile to be seen in the electron density.

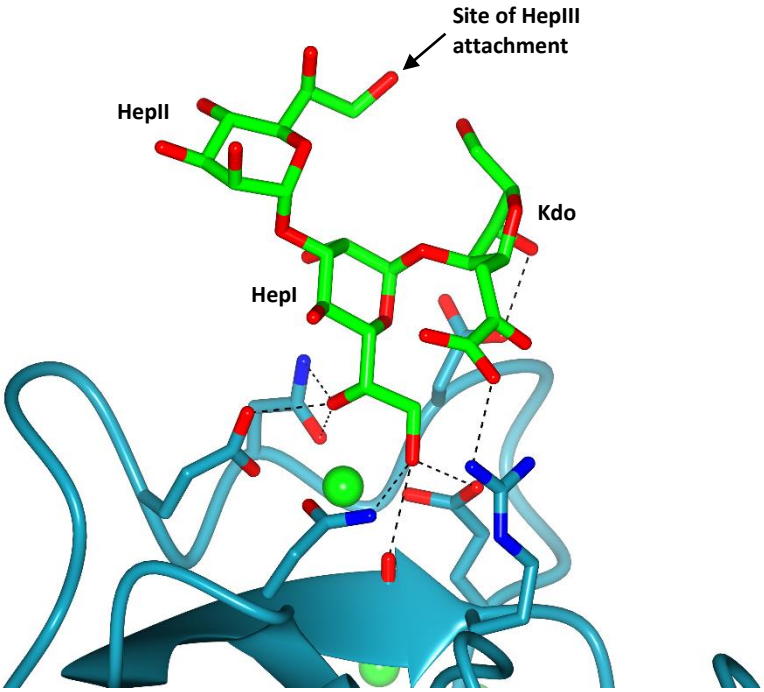


Figure 5-11 Site of attachment of HepIII to the O7 hydroxyl group of HepII. Both HepI and Kdo interact directly with the protein, while HepII interacts only with water molecules and residues of the symmetry-related trimer (not shown). Ligand coordinating residues are shown in cyan. Image generated using CCP4MG.

There is vast similarity between LPS inner-core structures of Gram-negative bacteria; and the ligand structure presented here is identical to LPS structures found in various other genera, such as:

- *Bordetella* – pertussis (Caroff & Karibian, 2000)
- *Burkholderia* – pneumonia (Ortega *et al.,* 2009)
- *E. coli* - gastroenterological diseases (Heinrichs *et al*., 1998)
- *Klebsiella* - pneumonia, septicaemia, meningitis (Vinogradov *et al.,* 2002)
- *Neisseria* - meningitis, gonorrhoea (Cox *et al.,* 2003)
- *Providencia* – UTIs (Kondakova *et al.*, 2006)
- *Pseudomonas* – sepsis (Bystrova *et al.,* 2004)
- *Serratia* – respiratory and urinary tract infections (Vinogradov *et al.,* 2003)
- *Shigella* – dysentery (Kondakova *et al.,* 2010)
- *Yersinia* – Bubonic plague (Vinogradov *et al.,* 2002)

Given this similarity, the binding mechanism presented here may be applicable when discussing the recognition of the many different bacteria by SP-D, and provides an understanding of how such a broad range of pathogens are bound – i.e. that despite vast variation in outer-core and O-antigen structures present on Gram-negative microbial surfaces, the inner core (and in particular, non-terminal HepI) is a conserved binding target for SP-D.

5.3.4 Anhydro-Kdo formation

Acid hydrolysis of intact LPS yields oligosaccharides with anhydro-Kdo on the reducing terminus rather than intact Kdo (Auzanneau *et al.,* 1991). Formation of anhydro-Kdo structures in *H. influenzae* lipopolysaccharides have been previously proposed when the phosphate moiety from C-4 of Kdo is β-eliminated (Danan *et al.,* 1982; Caroff *et al.,* 1987). Loss of phosphate from the 4-position of Kdo proceeds readily under these conditions because the C-4 substituent is β to the C-2 (anomeric) carbon. The reaction is believed to give rise to olefinic Kdo derivatives, which rearrange to form anhydro ring structures. This mechanism of elimination is thus only possible with *H. influenzae*-type lipid A structures that contain a phosphorylated single Kdo residue linked to lipid A. *E coli* and *Salmonella* LPS contains an unphosphorylated double-Kdo (Horstman *et al*., 2004) and so cannot undergo this modification. Electron density of the structure reveals that a five-, not six-membered, ring exists adjacent to the bound-heptose (see Fig 5-11). The density suggests that this Kdo residue also undergoes a structural change during mild acid hydrolysis. The mechanism of anhydro-Kdo formation is replicated below.

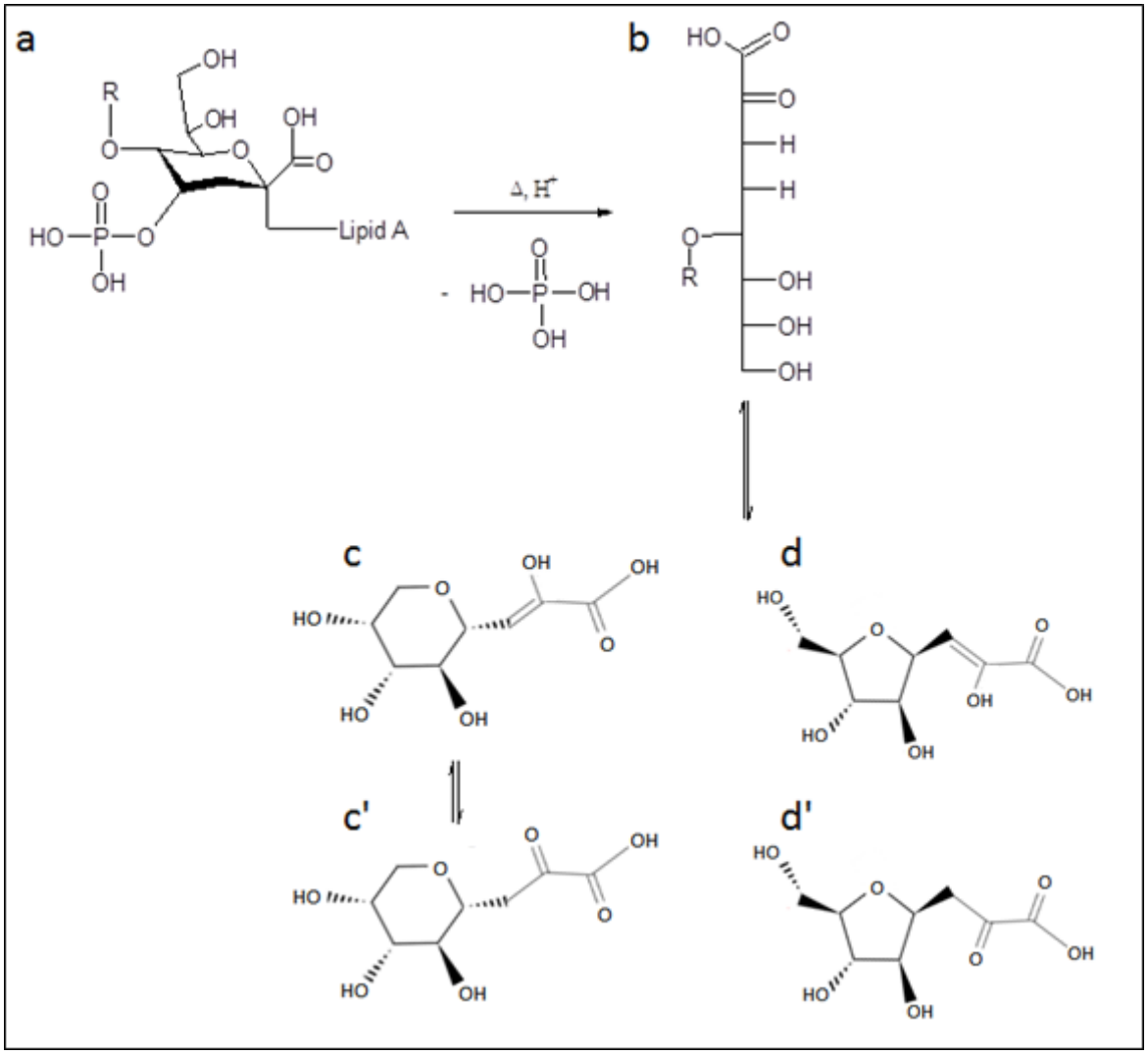


Figure 3-2 (replicated) Formation of the structures formed during mild acid hydrolysis of 3-Deoxy-D-manno-oct-2-ulosonic acid (Kdo, a); D-arabino-3-en-2-ulonic acid open-chain (b), the diasteromeric forms of the 4,8-anhydro (c and c') and the 4,7-anhydro (d and d') derivatives. R represents the extended inner core of LPS. Image adapted from Sioud *et al*., 2010.

Anhydro-Kdo residues (anKdo), rather than intact Kdo, as reducing terminal monosaccharides produced after acid hydrolysis have previously been determined for numerous lipopolysaccharides - and contain structures of either '*Salmonella*-type' with a Kdo at position C4, or '*Haemophilus*-type' with a phosphate group. Bacterial LPS species that display the former include *Aeromonas*, *Bordetella* and *Vibrionaceae* and produce core oligosaccharides that contain a mixture of diastereomeric 4,8- and 4,7-anhydro acids and an open-chain olefinic Kdo residue (Sioud *et al.,* 2010; Danan *et al*., 1982; Banoub *et al*., 2010). Species of bacteria with similar Kdo

arrangements to that of the tested *Salmonella* R7 strain include *Burkholderia, Chlamydia, Coxiella, E. coli, Klebsiella, Neisseria, Pseudomonas* and *Yersinia.* (Isshiki *et al*., 2003; Kosma *et al.,* 1988; Schramek *et al.,* 1985; Yethon *et al*., 2000; Fresno *et al*., 2006; Kahler, 2004; Bystrova *et al*., 2004; Vinogradov *et al.*, 2002). This is, however, the first instance of an anhydro-Kdo moiety from an LPS other than *H. influenzae* is seen in an electron density map. Although intact Kdo may assume a different position and orientation compared to that presented here, there will still be ample opportunity for interaction of Asp325 and Arg343 with the Kdo although these interactions may differ from those seen here - Asp325 with O6 hydroxyl group, and Arg343 with the tail group of the five-membered, anhydrous ring structure.

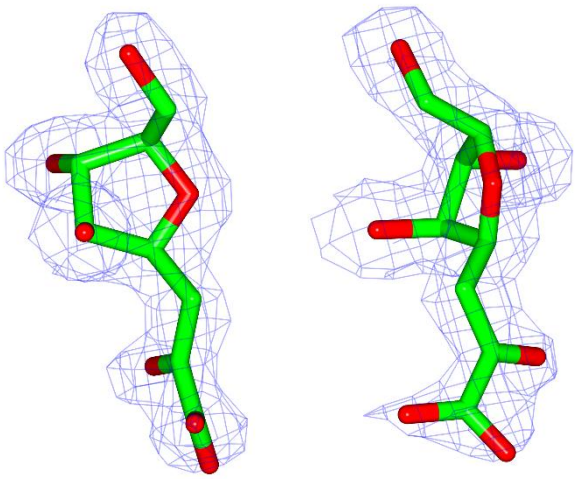


Figure 5-12 Electron density map of the anhydro-Kdo residue. There is clear evidence that this saccharide adopts a five-membered ring conformation, rather than the conventional six-membered structure. Image generated using CCP4MG.

6 Conclusions and future work

6.1 Conclusions

The role of SP-D in pathogen recognition involves binding of carbohydrate structures on the microbial surface, and a recombinant fragment containing just the head and neck regions has been proven therapeutically active (Sh rive *et al*., 2009). Crystals of this recombinant portion of SP-D were grown in an unconventional manner, by which calcium was added after the initial tray setup. Chosen crystals were flash-frozen after being soaked with a buffer containing MPD cryoprotectant and ligand, then the drop solution fully exchanged for 20% MPD cryobuffer. In the case of well CCS16A1, crystals were subjected to successive aliquots of 5-15% MPD and showed immediate degradation, and so were exchanged prior to the final additions of 20% cryobuffer. Three crystals were frozen and despite only soaking for 1 to 5 minutes, a data set was collected for crystal CCS16A11 which was processed at 1.77 Å and revealed a well-defined ligand in the binding pocket of chains B and C. Molecules packed in the crystal impose constraints on binding of longer structures in subunit A, and as such no ligand was visible in the electron density in this chain.

6.1.1 Investigating the binding of bacterial LPS by SP-D – recognition of heptose in the inner core.

This data provides insight in to the binding mechanism of bacterial lipopolysaccharide by SP-D, and the preferential binding to heptose 1 over heptose 2 in the inner core, as well as further interactions involving saccharides other than that bound at the Ca1 site.

In chains B and C, ligand binding was primarily observed via the glycerol sidechain hydroxyl groups of heptose 1, by amino acids in the CRD which also coordinate calcium ion Ca1; the OE2 atom of Glu321, ND2 of Asn323, OD1 atoms of Asn323 and Asn341 interact with the O6 hydroxyl group on the glycerol side chain, while the OE2 of Glu329, ND2 of Asn341 and the main-

chain carbonyl of Asp342 interact with the O7 group. The anhydro-Kdo monosaccharide also formed hydrogen bonds with residues Asp325 and Arg343, but no direct interaction was seen between the second, outer heptose and the protein. However, two direct hydrogen bonds and numerous hydrogen-bond bridges between HepII and the symmetry-related trimer were observed, as well as a single bond between saccharides of the ligand itself that vary between the B and C chain.

This structure confirms that SP-D binding of LPS and subsequent agglutination and clearance of Gram-negative bacteria is achieved through recognition and binding of inner core heptose residues. The inner core structure described here is conserved across a wide range of bacterial genera including *Bordetella*, *Burkholderia*, *E. coli*, *Haemophilus*, *Helicobacter, Klebsiella*, *Neisseria, Pasteurella, Providencia*, *Pseudomonas, Salmonella*, *Serratia*, *Shigella, Vibrio* and *Yersinia* species, and thus demonstrates the importance of SP-D as a pathogen-recognition molecule. A possible mechanism of immune system evasion is also suggested here, whereby lipopolysaccharide molecules effectively fold back on themselves to shield the inner core heptose residues that are typical targets of collectin binding. The structure also confirms the ability of SP-D to recognise non-terminal monosaccharides, which has previously not been shown structurally.

Overall, this work represents high-resolution structural insights into binding of bacterial lipopolysaccharides by surfactant protein-D, and provides a detailed binding mechanism and preference of the innermost heptose residue - as well as further interactions with the terminal anhydro-Kdo.

6.3 Future work

The ligand present in the binding site here consists of two heptose saccharides and an anhydro-Kdo residue resulting from the mild acis hydrolysis of the intact LPS. While this does not represent the structure of the R7 LPS *in vivo* due to the presence of the anhydro KdO, it provides valuable insights and data on the recognition of the triheptosyl inner core of a range of bacterial targets of

SP-D. This may be enhanced by the presence of an intact Kdo rather than the 5-membered anhydro Kdo although it seems likely that the major binding determinant is the inner heptose HepI as demonstrated here. The delipidation procedure of Phillips *et al.,* 1992 does not involve beta-elimination of phosphate from the Kdo and subsequent rearrangement to a 5-membered ring, and as such is capable of providing the true physiological ligand. The indication from the structure presented here is that the crystal would allow binding of this ligand in 2 of the three chains, and that interaction with Arg343 would form part of the binding mechanism.

In addition to the small ligands and monosaccharides previously crystallised in complex with SP-D, crystal structures of bound pathogen surface molecules such as fucose glycoconjugates, lipoarabinomannan and α1-2-linked trimannose are yet to be elucidated. Revealing the mechanism by which these ligands are bound may provide a deeper understanding of SP-D neutralisation of schistosomes, mycobacteria and influenza A viruses, respectively. Given the relative sizes of these example ligands, they are also small enough to be soaked in to crystals of the physiologically relevant, biologically and therapeutically active recombinant protein used in the present study.

Detailed insights such as those presented here into exactly how SP-D binds pathogens suggest routes towards the design of recombinant fragments with modified binding pockets (and flanking residues) capable of enhanced or altered specificity for bacterial and viral ligands. Designer recombinant fragments targeted at specific diseases offer a potential route to the treatment of disease and infection. An example pf this is provided by Nikolaidis *et al*., 2014, where mutation of SP-D residues Asp325 and Arg343, as D325A+R343V and D325S+R343V, showed a marked increase in SP-D antiviral activity against seasonal strains of influenza A virus. A wealth of literature confirms that these binding-pocket flanking residues are a key part of both recognition and function.

An interesting and well-researched area of SP-D function is its ability to recognise and bind to receptors to stimulate an immune response. There are however, no structural studies of SP-D in complex with molecules other than small ligands, which is in part responsible for the incomplete understanding of the basis of receptor specificity and of the activation of immune mechanisms by SP-D. It may be possible that SP-D utilises the same site, but at distant locations on the physiological dodecameric cruciform SP-D molecule, for both bacterial and receptor binding. This possiblility is suggested by the calcium-dependent binding of the receptor SIRPα (Fournier *et al*., 2012). While X-ray structural studies are inherently unable to show such a scenario due to the nature of the physiological SP-D molecule, structural studies aimed at revealing a receptor fragment bound by the SP-D CRD would be a significant step forward in our understanding.

It may, however, be that binding of a physiological ligand is required to reveal the receptor binding mechanism, consistent with another key molecule in innate immunity, C-reactive protein, which appears to be unable to recognise receptors or activate complement unless a true physiological ligand is also bound (Agrawal *et al.*, 2001). This presents a significant challenge. Park *et al*., (2009), successfully cocrystallised intact, rough lipopolysaccharide bound to a TLR4-MD2 complex - a feat accomplished by sonication of an LPS solution, followed by incubation with the protein mixture. Repeating this experiment using SP-D and a relevant intact LPS could shed light on the mechanism by which an immune response is provoked once presented with a microbe.

More generally, the roles of SP-D include far more than pathogen recognition; for instance, surfactant therapy has been highly successful in reducing mortality from neonatal respiratory distress syndrome, commonly seen in premature infants born less than 28 weeks' gestation (Clark, 2010) and from meconium aspiration syndrome (Dargville & Mills, 2005). Current surfactant therapies do not include SP-A or SP-D, but considering the roles of reducing lung inflammation, surfactant homeostasis and lipid regulation, inclusion of recombinant SP-D fragments in this

treatment may help offset the risk of development of respiratory distress and chronic lung disease.

The research presented in this thesis is a major advance towards the definition of the recognition properties of surfactant protein D, a key element of innate immunity and front-line host defence. It represents a further and significant step towards the exploitation of the biologically and therapeutically recombinant fragment in the prevention, management and treatment of a variety of respiratory and lung diseases.

References

1. Aduse-Opoku, J., Slaney, J., Hashim, A., Gallagher, A., Gallagher, R., Rangarajan, M., Boutaga, K., Laine, M., Van Winkelhoff, A. and Curtis, M. (2005). Identification and Characterization of the Capsular Polysaccharide (K-Antigen) Locus of Porphyromonas gingivalis. Infection and Immunity, 74(1), pp.449-460.
2. Agrawal, A., Shrive, A., Greenhough, T. and Volanakis, J. (2001). Topology and Structure of the C1q-Binding Site On C-Reactive Protein. The Journal of Immunology, 166(6), pp.3998-4004.
3. Alexander, C. and Rietschel, E. (2001). Invited review: Bacterial lipopolysaccharides and innate immunity. Journal of Endotoxin Research, 7(3), pp.167-202.
4. Allen, M. (2004). Arg343 in human surfactant protein D governs discrimination between glucose and N-acetylglucosamine ligands. Glycobiology, 14(8), pp.693-700.
5. Allen, M., Laederach, A., Reilly, P. and Mason, R. (2001a). Polysaccharide Recognition by Surfactant Protein D: Novel Interactions of a C-Type Lectin with Nonterminal Glucosyl Residues. *Biochemistry*, 40(26), pp.7789-7798.
6. Allen, M., Voelker, D. and Mason, R. (2001b). Interactions of Surfactant Proteins A and D with *Saccharomyces cerevisiae* and *Aspergillus fumigatus*. *Infection and Immunity*, 69(4), pp.2037-2044.
7. Ampel, N., Nelson, D., Li, L., Dionne, S., Lake, D., Simmons, K. and Pappagianis, D. (2005). The Mannose Receptor Mediates the Cellular Immune Response in Human Coccidioidomycosis. *Infection and Immunity*, 73(4), pp.2554-2555.
8. Appelmelk, B., Negrini, R., Moran, A. and Kuipers, E. (1997). Molecular mimicry between *Helicobacter pylori* and the host. *Trends in Microbiology*, 5(2), pp.70-73.
9. Asherie, N. (2012). A Dialogue about Protein Crystallization and Phase Diagrams. *Protein & Peptide Letters*, 19(7), pp.708-713.

10. Aurell, C. and Wistrom, A. (1998). Critical Aggregation Concentrations of Gram-Negative Bacterial Lipopolysaccharides (LPS). Biochemical and Biophysical Research Communications, 253(1), pp.119-123.
11. Awasthi, S., Magee, D. and Coalson, J. (2004). *Coccidioides posadasii* infection alters the expression of pulmonary surfactant proteins (SP)-A and SP-D. *Respiratory Research*, 5(1), pp.28-33
12. Axelgaard, E., Jensen, L., Dyrlund, T., Nielsen, H., Enghild, J., Thiel, S. and Jensenius, J. (2013). Investigations on Collectin Liver 1. *Journal of Biological Chemistry*, 288(32), pp.23407-23420.
13. Backhed, F., Soderhall, M., Ekman, P., Normark, S. and Richter-Dahlfors, A. (2001). Induction of innate immune responses by Escherichia coli and purified lipopolysaccharide correlate with organ- and cell-specific expression of Toll-like receptors within the human urinary tract. Cell Microbiol, 3(3), pp.153-158.
14. Banoub, J., Aneed, A., Cohen, A. and Joly, N. (2010). Structural investigation of bacterial lipopolysaccharides by mass spectrometry and tandem mass spectrometry. Mass Spectrom. Rev., 29(4), pp.606-650.
15. Barlo, N., Van Moorsel, C., Ruven, H., Zanen, P., Van den Bosch, J. and Grutters, J. (2009). Surfactant protein-D predicts survival in patients with idiopathic pulmonary fibrosis. *Sarcoidosis Vasc. Diffuse Lung Dis*., 26(2), pp.155-61.
16. Battye, T., Kontogiannis, L., Johnson, O., Powell, H. and Leslie, A. (2011) Imosflm : A New Graphical Interface For Diffraction-Image Processing With MOSFLM. Acta Crystallogr D Biol Cryst., 67(4), pp.271-281.
17. Bayer, H., Bonnar, J., Phizackerley, P., Moore, R. and Wylie, F. (1973). Amniotic fluid phospholipids in normal and abnormal pregnancy. BJOG: An International Journal of Obstetrics and Gynaecology, 80(4), pp.333-337.
18. Beharka, A., Gaynor, C., Kang, B., Voelker, D., McCormack, F. and Schlesinger, L. (2002). Pulmonary Surfactant Protein A Up-Regulates Activity of the Mannose Receptor, a Pattern

Recognition Receptor Expressed on Human Macrophages. *The Journal of Immunology*, 169(7), pp.3565-3573.

19. Bianchi, M. (2006). DAMPs, PAMPs and alarmins: all we need to know about danger. *Journal of Leukocyte Biology*, 81(1), pp.1-5.
20. Blow, D., Chayen, N., Lloyd, L. and Saridakis, E. (1994). Control of nucleation of protein crystals. Protein Science, 3(10), pp.1638-1643.
21. Boettner, D., Huston, C. and Petri, W. (2002). Galactose/N-acetylgalactosamine lectin: The coordinator of host cell killing. *J. Biosci.*, 27(6), pp.553-557.
22. Bonella, F., Volpe, A., Caramschi, P., Nava, C., Ferrari, P., Schenk, K., Ohshimo, S., Costabel, U. and Ferrari, M. (2011). Surfactant protein D and KL-6 serum levels in systemic sclerosis: correlation with lung and systemic involvement. *Sarcoidosis Vasc. Diffuse Lung Dis*., 28, pp.27-33.
23. Borisenko, E. and Kolesnikov, N. (2012). Crystallization and Materials Science of Modern Artificial and Natural Crystals. Rijeka: InTech, pp.249-276.
24. Borron, P., Crouch, E., Lewis, J., Wright, J., Possmayer, F. and Fraher, L. (1998) Recombinant rat surfactant-associated protein D inhibits human T lymphocyte proliferation and IL-2 production. *J. Immunol*. 161(4), pp.599–4603.
25. Botas, C., Poulain, F., Akiyama, J., Brown, C., Allen, L., Goerke, J., Clements, J., Carlson, E., Gillespie, A., Epstein, C. and Hawgood, S. (1998). Altered Surfactant Homeostasis and Alveolar Type II Cell Morphology in Mice Lacking Surfactant Protein D. *Proceedings of the National Academy of Sciences* 95(20), pp.11869-11874.
26. Brade, H. (1999). Endotoxin in health and disease. New York: Marcel Dekker.
27. Brown, A., Quinn, M., Karim, Z., Conaghan, P., Peterfy, C., Hensor, E., Wakefield, R., O'Connor, P. and Emery, P. (2006). Presence of significant synovitis in rheumatoid arthritis patients with disease-modifying antirheumatic drug–induced clinical remission: Evidence from an imaging study may explain structural progression. Arthritis & Rheumatism, 54(12), pp.3761-3773.

28. Brown-Augsburger, P., Chang, D., Rust, K. and Crouch, E. (1996). Biosynthesis of Surfactant Protein D: Contributions of conserved NH2-terminal cysteine residues and collagen helix formation to assembly and secretion. *Journal of Biological Chemistry*, 271(31), pp.18912-18919.

29. Bufler, P., Schmidt, B., Schikor, D., Bauernfeind, A., Crouch, E. and Griese, M. (2003). Surfactant Protein A and D Differently Regulate the Immune Response to Nonmucoid Pseudomonas aeruginosa and Its Lipopolysaccharide. Am. J. Respir. Cell Mol. Biol., 28(2), pp.249-256.

30. Bystrova, O., Lindner, B., Moll, H., Kocharova, N., Knirel, Y., Zahringer, U. and Pier, G. (2004). Full Structure of the Lipopolysaccharide of Pseudomonas aeruginosa Immunotype 5. Biochemistry (Moscow), 69(2), pp.170-175

31. Caroff, M. and Karibian, D. (2003). Structure of bacterial lipopolysaccharides. Carbohydrate Research, 338(23), pp.2431-2447.

32. cdc.gov, (2015). CDC - Gram-negative Bacteria Infections in Healthcare Settings - HAI. [online] Available at: http://www.cdc.gov/hai/organisms/gram-negative-bacteria.html [Accessed 12 Apr. 2015].

33. Chen, V., Arendall, W., Headd, J., Keedy, D., Immormino, R., Kapral, G., Murray, L., Richardson, J. and Richardson, D. (2009). MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst Sect D*, 66(1), pp.12-21.

34. Chiba, H. (2003). Pulmonary Surfactant Proteins A and D Recognize Lipid Ligands on Mycoplasma pneumoniae and Markedly Augment the Innate Immune Response to the Organism. Chest, 123(90030), pp.426S-a-426.

35. Chiba, H., Pattanajitvilai, S., Evans, A., Harbeck, R. and Voelker, D. (2002). Human Surfactant Protein D (SP-D) Binds *Mycoplasma pneumoniae* by High Affinity Interactions with Lipids. *Journal of Biological Chemistry*, 277(23), pp.20379-20385.

36. Chida, S., Fujiwara, T., Konishi, M., Shimada, S. and Takahashi, A. (1997). Surfactant proteins and stable microbubbles in tracheal aspirates of infants with respiratory distress syndrome: relation to the degree of respiratory failure and response to exogenous surfactant. European Journal of Paediatrics, 156(2), pp.131-138.

37. Clark, H., Mackay, R., Deadman, M., Hood, D., Moxon, D., Townsend, J., Reid, K., Ahmed, A., Shaw, A., Greenhough, T. and Shrive, A. *Haemophilus influenzae* evades innate immune defence mediated by surfactant protein D by blocking multiple interactions with the LPS core. Unpublished.

38. Clark, Howard W. (2010). Untapped Therapeutic Potential Of Surfactant Proteins: Is There A Case For Recombinant SP-D Supplementation In Neonatal Lung Disease?. Neonatology 97(4), pp.380-387.

39. Cooley, J., McDonald, B., Accurso, F., Crouch, E. and Remold-O'Donnell, E. (2008). Patterns of neutrophil serine protease-dependent cleavage of surfactant protein D in inflammatory lung disease. *Journal of Leukocyte Biology*, 83(4), pp.946-955.

40. Cowtan, K. (1994). Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography, 31, pp.34-38

41. Cox, A., Wright, J., Gidney, M., Lacelle, S., Plested, J., Martin, A., Moxon, E. and Richards, J. (2003). Identification of a novel inner-core oligosaccharide structure in Neisseria meningitidis lipopolysaccharide. Eur J Biochem, 270(8), pp.1759-1766

42. Crabtree, J. E., and R. A. Wilson. (1986). The role of pulmonary cellular reactions in the resistance of vaccinated mice to Schistosoma mansoni. Parasite Immunol. 8(3), pp.265–285.

43. Crouch, E., Chang, D., Rust, K., Persson, A. and Heuser, J. (1994). Recombinant Pulmonary Surfactant Protein D: post-translational modification and molecular assembly. *Journal of Biological Chemistry*, 269(22), pp.15808-15813.

44. Crouch, E., Hartshorn, K., Horlacher, T., McDonald, B., Smith, K., Cafarella, T., Seaton, B., Seeberger, P. and Head, J. (2009). Recognition of Mannosylated Ligands and Influenza A Virus

by Human Surfactant Protein D: Contributions of an Extended Site and Residue 343. *Biochemistry*, 48(15), pp.3335-3345.

45. Crouch, E., McDonald, B., Smith, K., Cafarella, T., Seaton, B. and Head, J. (2006). Contributions of Phenylalanine 335 to Ligand Recognition by Human Surfactant Protein D: Ring interactions with SP-D ligands. *Journal of Biological Chemistry*, 281(26), pp.18008-18014.
46. Crouch, E., McDonald, B., Smith, K., Roberts, M., Mealy, T., Seaton, B. and Head, J. (2007). Critical Role of Arg/Lys343 in the Species-Dependent Recognition of Phosphatidylinositol by Pulmonary Surfactant Protein D. *Biochemistry*, 46(17), pp.5160-5169.
47. Crouch, E., Rust, K., Veile, R., Donis-Keller, H. and Grosso, L. (1993). Genomic organization of human surfactant protein D (SP-D). SP-D is encoded on chromosome 10q22.2-23.1. *Biochemistry*, 268(4), pp.2976-2983.
48. Currie, A., Stewart, G. and McWilliam, A. (2000). Alveolar Macrophages Bind and Phagocytose Allergen- Containing Pollen Starch Granules Via C-Type Lectin and Integrin Receptors: Implications for Airway Inflammatory Disease. The Journal of Immunology, 164(7), pp.3878-3886.
49. Dahms, N. (2002). P-type lectins. *Biochimica et Biophysica Acta - General Subjects*, 1572(2-3), pp.317-340.
50. Dargaville, P. and Mills, J. (2005). Surfactant Therapy For Meconium Aspiration Syndrome. Drugs, 65(18), pp.2569-2591.
51. Danan, A., Mondange, M., Sarfati, S. and Szab, P. (1982). Synthesis and behaviour under acidic conditions of 2-deoxy-D-arabino-hexopyranose and 3-deoxy-2-ketoaldonic acids bearing O-phosphono or O-glucosyl substituents at position β to the carbonyl function. J. Chem. Soc., Perkin Trans. 1, p.1275.
52. DeLucia, A., Six, D., Caughlan, R., Gee, P., Hunt, I., Lam, J. and Dean, C. (2011). Lipopolysaccharide (LPS) Inner-Core Phosphates Are Required for Complete LPS Synthesis and

Transport to the Outer Membrane in Pseudomonas aeruginosa PAO1. mBio., 2(4), pp.e00142-11.

53. Derewenda, Z. and Vekilov, P. (2005). Entropy and surface engineering in protein crystallization. *Acta Cryst Sect D*, 62(1), pp.116-124.
54. Dessau, M. and Modis, Y. (2011). Protein Crystallization for X-ray Crystallography. *Journal of Visualized Experiments*, (47).
55. diamond.ac.uk. Anon, (2015). *Diamond Light Source - Science - Injection system and linac*. [online] Available at: http://www.diamond.ac.uk/Science/Machine/Components/linac. [Accessed 5 Apr. 2015].
56. Dinarillo, C. (1991). Interleukin-1 and interleukin-1 antagonism. *Blood*, 77(8), pp.1627-162.
57. Drickamer, K. (1992). Engineering galactose-binding activity into a C-type mannose-binding protein. *Nature*, 360(6400), pp.183-186.
58. Drickamer, K. and Taylor, M. (1993). Biology of Animal Lectins. *Annual Review of Cell Biology*, 9(1), pp.237-264.
59. Ducruix, A. and Giegé, R. (1999). Crystallization of nucleic acids and proteins. Oxford: Oxford University Press.
60. Eddie Ip, W., Takahashi, K., Alan Ezekowitz, R. and Stuart, L. (2009). Mannose-binding lectin and innate immunity. *Immunological Reviews*, 230(1), pp.9-21.
61. Emsley, P. and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Cryst Sect D*, 60(12), pp.2126-2132.
62. Emsley, P., Lohkamp, B., Scott, W. and Cowtan, K. (2010). Features and development of Coot. *Acta Cryst Sect D*, 66(4), pp.486-501.
63. Englemann, B. and Massberg, S. (2014). Innate Immunity, Coagulation, and Thrombosis. *Blood*, 124(21).
64. Erpenbeck, V. (2004). Surfactant protein D increases phagocytosis and aggregation of pollen-allergen starch granules. AJP: Lung Cellular and Molecular Physiology, 288(4), pp.L692-L698.

65. Erridge, C., Bennett-Guerro, E. and Poxton, I. (2002). Structure and function of lipopolysaccharide. Microbes and Infection, 4(8), pp.837-851.

66. Ferguson J., Voelker, D., McCormack, F. and Schlesinger, L. (1999). Surfactant protein D binds to *Mycobacterium tuberculosis* bacilli and Lipoarabinomannan via carbohydrate–lectin interactions resulting in reduced phagocytosis of the bacteria by macrophages. *J. Immunol*. 163, pp.312–321.

67. Floros, J., Lin, H., García, A., Salazar, M., Guo, X., DiAngelo, S., Montaño, M., Luo, J., Pardo, A. and Selman, M. (2000). Surfactant Protein Genetic Marker Alleles Identify a Subgroup of Tuberculosis in a Mexican Population. The Journal of Infectious Diseases, 182(5), pp.1473-1478.

68. Fournier, B. Andargachev, R., Robin, A., Laur, O., Voelker, D., Lee, W., Weber, D. and Parkos, C. Surfactant Protein D (Sp-D) Binds To Membrane-Proximal Domain (D3) Of Signal Regulatory Protein (SIRP), A Site Distant From Binding Domain Of CD47, While Also Binding To Analogous Region On Signal Regulatory Protein (SIRP). *Journal of Biological Chemistry* 287(23), pp.19386-19398.

69. Fresno, S., Jimenez, N., Canals, R., Merino, S., Corsaro, M., Lanzetta, R., Parrilli, M., Pieretti, G., Regue, M. and Tomas, J. (2006). A Second Galacturonic Acid Transferase Is Required for Core Lipopolysaccharide Biosynthesis and Complete Capsule Association with the Cell Surface in Klebsiella pneumoniae. Journal of Bacteriology, 189(3), pp.1128-1137.

70. Fritz, J., Ferrero, R., Philpott, D. and Girardin, S. (2006). Nod-like proteins in immunity, inflammation and disease. *Nat Immunol*, 7(12), pp.1250-1257.

71. Galloway, S. and Raetz, C. (1990). A mutant of *Escherichia coli* defective in the first step of endotoxin biosynthesis. *Journal of Biological Chemistry*, 165(11), pp.6394-69402.

72. Gardai, S., Xiao, Y., Dickinson, M., Nick, J., Voelker, D., Greene, K. and Henson, P. (2003). By Binding SIRPα or Calreticulin/CD91, Lung Collectins Act as Dual Function Surveillance Molecules to Suppress or Enhance Inflammation. Cell, 115(1), pp.13-23.

73. Gauglitz, G., Callenberg, H., Weindl, G. and Korting, H. (2012). Host Defence against Candida albicans and the Role of Pattern-recognition Receptors. Acta Dermato Venereologica, 92(3), pp.291-298.

74. Gaunsbaek, M., Rasmussen, K., Beers, M., Atochina-Vasserman, E. and Hansen, S. (2013). Lung Surfactant Protein D (SP-D) Response and Regulation during Acute and Chronic Lung Injury. Lung, 191(3), pp.295-303.

75. Gemou-Engesaeth, V., Laliotou, N., Corrigan, C., Chrousos, G. and Haczku, A. (2014). Expression of pulmonary surfactant protein D (SP-D) and interleukin 13 in the serum of atopic and non-atopic severe paediatric asthmatics: effects of glucocorticoid and sodium cromoglycate treatment. Clinical and Translational Allergy, 4(1), pp.3

76. Gill, S. and von Hippel, P. (1989). Calculation of protein extinction coefficients from amino acid sequence data. Analytical Biochemistry, 182(2), pp.319-326.

77. Goh, B., Rynkiewicz, M., Cafarella, T., White, M., Hartshorn, K., Allen, K., Crouch, E., Calin, O., Seeberger, P., Schulten, K. and Seaton, B. (2013). Molecular Mechanisms of Inhibition of Influenza by Surfactant Protein D Revealed by Large-Scale Molecular Dynamics Simulation. *Biochemistry*, 52(47), pp.8527-8538.

78. Greene, K., King, T., Kuroki, Y., Bucher-Bartelson, B., Hunninghake, G., Newman, L., Nagae, H. and Mason, R. (2002). Serum surfactant proteins-A and -D as biomarkers in idiopathic pulmonary fibrosis. European Respiratory Journal, 19(3), pp.439-446.

79. Greene, K., Kuroki, Y., Bucher Bartelson, B., King, T., Hunninghake, G., Parsons, P., Newman, L., Nagoe, H., Tudor, R., Kuhn, C. and Mason, R. (2001). Serum Concentrations of Surfactant Proteins A and D Predict Mortality in Patients With Idiopathic Pulmonary Fibrosis. Chest, 120.

80. Greene, K., Wright, J., Steinberg, K., Ruzinski, J., Caldwell, E., Wong, W., Hull, W., Whitsett, J., Akino, T., Kuroki, Y., Nagae, J., Hudson, L. and Martin, T. (1999). Serial Changes in Surfactant-associated Proteins in Lung and Serum before and after Onset of ARDS. Am. J. Respir. Crit. Care Med., 160(6), pp.1843-1850.

81. Griese, M., Essl, R., Schmidt, R., Rietschel, E., Ratjen, F., Ballmann, M. and Paul, K. (2004). Pulmonary Surfactant, Lung Function, and Endobronchial Inflammation in Cystic Fibrosis. Am. J. Respir. Crit. Care Med., 170(9), pp.1000-1005.

82. Guo, C., Atochina-Vasserman, E., Abramova, E., Foley, J., Zaman, A., Crouch, E., Beers, M., Savani, R. and Gow, A. (2008). S-Nitrosylation of Surfactant Protein-D Controls Inflammatory Function. Plos Biol, 6(11), p.e266.

83. Guo, X., Lin, H., Lin, Z., Montaño, M., Sansores, R., Wang, G., Diangelo, S., Pardo, A., Selman, M. and Floros, J. (2001). Surfactant protein gene A, B, and D marker alleles in chronic obstructive pulmonary disease of a Mexican population. European Respiratory Journal, 18(3), pp.482-490.

84. Haczku, A. (2008). Protective role of the lung collectins surfactant protein A and surfactant protein D in airway inflammation. *Journal of Allergy and Clinical Immunology*, 122(5), pp.861-879.

85. Haczku, A., Cao, Y., Vass, G., Kierstein, S., Nath, P., Atochina-Vasserman, E., Scanlon, S., Li, L., Griswold, D., Chung, K., Poulain, F., Hawgood, S., Beers, M. and Crouch, E. (2006). IL-4 and IL-13 Form a Negative Feedback Circuit with Surfactant Protein-D in the Allergic Airway Response. The Journal of Immunology, 176(6), pp.3557-3565.

86. Hajela, K. (2002). The Biological Functions of MBL-Associated Serine Proteases (MASPs). *Immunobiology*, 205(4-5), pp.467-475.

87. Håkansson, K. and Reid, K. (2000). Collectin structure: A review. *Protein Science*, 9(9), pp.1607-1617.

88. Hartshorn, K., Crouch, E., White, M., Colamussi, M., Kakkanatt, A., Tauber, B., Shepherd, V., and Sastry, K. (1998). Pulmonary surfactant proteins A and D enhance neutrophil uptake of bacteria. *Am.J.Physiol*. 274, pp.L958–L969.

89. Hartshorn, K., Webby, R., White, M., Tecle, T., Pan, C., Boucher, S., Moreland, R., Crouch, E. and Scheule, R. (2008). Role of viral hemagglutinin glycosylation in anti-influenza activities of recombinant surfactant protein D. Respiratory Research, 9(1), p.65.

90. Hartshorn, K., White, M., Tecle, T., Holmskov, U. and Crouch, E. (2006). Innate Defense against Influenza A Virus: Activity of Human Neutrophil Defensins and Interactions of Defensins with Surfactant Protein D. *Journal of Immunology*, 176(11), pp.6962-6972.

91. Hartshorn, K., White, M., Voelker, D., Coburn, J., Zaner, K. and Crouch, E. (2000). Mechanism of binding of surfactant protein D to influenza A viruses: importance of binding to haemagglutinin to antiviral activity. *Biochemistry*, 351(2), pp.449-458.

92. Hassell, A., An, G., Bledsoe, R., Bynum, J., Carter, H., Deng, S., Gampe, R., Grisard, T., Madauss, K., Nolte, R., Rocque, W., Wang, L., Weaver, K., Williams, S., Wisely, G., Xu, R. and Shewchuk, L. (2006). Crystallization of protein–ligand complexes. *Acta Cryst Sect D*, 63(1), pp.72-79.

93. Heinrichs, D., Yethon, J. and Whitfield, C. (1998). Molecular basis for structural diversity in the core regions of the lipopolysaccharides of Escherichia coli and Salmonella enterica. Molecular Microbiology, 30(2), pp.221-232.

94. Helander, I., Kato, Y., Kilpelainen, I., Kostiainen, R., Lindner, B., Nummila, K., Sugiyama, T. and Yokochi, T. (1996). Characterization of Lipopolysaccharides of Polymyxin-Resistant and Polymyxin-Sensitive Klebsiella pneumoniae O3. Eur J Biochem, 237(1), pp.272-278.

95. Hickling, T., Bright, H., Wing, K., Gower, D., Martin, S., Sim, R. and Malhotra, R. (1999). A recombinant trimeric surfactant protein D carbohydrate recognition domain inhibits respiratory syncytial virus infection *in vitro* and *in vivo*. *European Journal of Immunology*, 29(11), pp.3478-3484.

96. Hills, B. (1994). Further studies of the role of surfactant in premature rupture of the membranes. American Journal of Obstetrics and Gynecology, 170(1), pp.195-201.

97. Hoare, A., Bittner, M., Carter, J., Alvarez, S., Zaldivar, M., Bravo, D., Valvano, M. and Contreras, I. (2006). The Outer Core Lipopolysaccharide of Salmonella enterica Serovar Typhi

Is Required for Bacterial Entry into Epithelial Cells. Infection and Immunity, 74(3), pp.1555-1564.

98. Hohwy, T., Otkjaer, K., Madsen, J., Søerensen, G., Nielsen, O., Vestergaard, C., Steiniche, T., Holmskov, U. and Lomholt, H. (2006). Surfactant protein D in atopic dermatitis and psoriasis. Exp Dermatol, 15(3), pp.168-174.

99. Holgate, S. (1999). Air pollution and health. San Diego: Academic Press.

100. Holmskov, U., Fischer, P., Rothmann, A. and Højrup, P. (1996). Affinity and kinetic analysis of the bovine plasma C-type lectin collectin-43 (CL-43) interacting with mannan. *FEBS Letters*, 393(2-3), pp.314-316.

101. Holst, O. (2011). Structure of the Lipopolysaccharide Core Region. Bacterial Lipopolysaccharides, pp.21-39.

102. Honda, Y. (1996). Decreased Contents of Surfactant Proteins A and D in BAL Fluids of Healthy Smokers. CHEST, 109(4), p.1006.

103. Ihn, H., Asano, Y., Kubo, M., Yamane, K., Jinnin, M., Yazawa, N., Fujimoto, M. and Tamaki, K. (2002). Clinical significance of serum surfactant protein D (SP-D) in patients with polymyositis/dermatomyositis: correlation with interstitial lung disease. Rheumatology, 41(11), pp.1268-1272.

104. Isshiki, Y., Zähringer, U. and Kawahara, K. (2003). Structure of the core-oligosaccharide with a characteristic d-glycero-α-d-talo-oct-2-ulosylonate-(2→4)-3-deoxy-d-manno-oct-2-ulosonate [α-Ko-(2→4)-Kdo] disaccharide in the lipopolysaccharide from Burkholderia cepacia. Carbohydrate Research, 338(23), pp.2659-2666.

105. Janeway, C. (2001). *Immunobiology 5*. New York: Garland Pub.

106. Jang, S., Ohtani, K., Fukuoh, A., Yoshizaki, T., Fukuda, M., Motomura, W., Mori, K., Fukuzawa, J., Kitamoto, N., Yoshida, I., Suzuki, Y. and Wakamiya, N. (2008). Scavenger Receptor Collectin Placenta 1 (CL-P1) Predominantly Mediates Zymosan Phagocytosis by Human Vascular Endothelial Cells. *Journal of Biological Chemistry*, 284(6), pp.3956-3965.

107. Jones, T., Zou, J., Cowan, S. and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst Sect A*, 47(2), pp.110-119.

108. Jounblat, R., Kadioglu, A., Iannelli, F., Pozzi, G., Eggleton, P. and Andrew, P. (2004). Binding and Agglutination of *Streptococcus pneumoniae* by Human Surfactant Protein D (SP-D) Vary between Strains, but SP-D Fails To Enhance Killing by Neutrophils. *Infection and Immunity*, 72(2), pp.709-716.

109. Kahler, C. (2004). Inner core assembly and structure of the lipooligosaccharide of Neisseria meningitidis: capacity of strain NMB to express all known immunotype epitopes. Glycobiology, 15(4), pp.409-419.

110. Kankavi, O. (2006). Increased Expression of Surfactant Protein A and D in Rheumatoid Arthritic Synovial Fluid. *Croat. Med. J.,* 47(1), pp.155-161.

111. Kawahara, K., Seydel, U., Matsuura, M., Danbara, H., Rietschel, E. and Zähringer, U. (1991). Chemical structure of glycosphingolipids isolated from Sphingomonas paucimobilis. FEBS Letters, 292(1-2), pp.107-110.

112. Khamri, W., Moran, A., Worku, M., Karim, Q., Walker, M., Annuk, H., Ferris, J., Appelmelk, B., Eggleton, P., Reid, K. and Thursz, M. (2005). Variations in *Helicobacter pylori* Lipopolysaccharide to Evade the Innate Immune Component Surfactant Protein D. *Infection and Immunity*, 73(11), pp.7677-7686.

113. Kishore, U., Greenhough, T., Waters, P., Shrive, A., Ghai, R., Kamran, M., Bernal, A., Reid, K., Madan, T. and Chakraborty, T. (2006). Surfactant proteins SP-A and SP-D: Structure, function and receptors. *Molecular Immunology*, 43(9), pp.1293-1315.

114. Kitaichi, N., Kitamura, M., Namba, K., Ishida, S. and Ohno, S. (2010). Elevation of surfactant protein D, a pulmonary disease biomarker, in the sera of uveitis patients with sarcoidosis. Japanese Journal of Ophthalmology, 54(1), pp.81-84.

115. Kleywegt, G. (2000). Validation of protein crystal structures. *Acta Cryst Sect D*, 56(3), pp.249-265.

116. Kleywegt, G. and Jones, T. (1996). Phi/Psi-chology: Ramachandran revisited. *Structure*, 4(12), pp.1395-1400.

117. Kollar, R., Reinhold, B., Petrakova, E., Yeh, H., Ashwell, G., Drgonova, J., Kapteyn, J., Klis, F. and Cabib, E. (1997). Architecture of the Yeast Cell Wall: Beta(1->6)-glucan interconnects mannoprotein, Beta(1->3)-glucan and chitin. Journal of Biological Chemistry, 272(28), pp.17762-17775.

118. Kondakova, A., Vinogradov, E., Lindner, B., Kocharova, N., Rozalski, A. and Knirel, Y. (2006). Elucidation of the Lipopolysaccharide Core Structures of Bacteria of the Genus Providencia. Journal of Carbohydrate Chemistry, 25(6), pp.499-520.

119. Knudsen, L., Wucherpfennig, K., Mackay, R., Townsend, P., Mühlfeld, C., Richter, J., Hawgood, S., Reid, K., Clark, H. and Ochs, M. (2009). A Recombinant Fragment of Human Surfactant Protein D Lacking the Short Collagen-Like Stalk Fails to Correct Morphological Alterations in Lungs of SP-D Deficient Mice. The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology, 292(2), pp.183-189.

120. Kondakova, A., Vinogradov, E., Shekht, M., Markina, A., Lindner, B., L'vov, V., Aparin, P. and Knirel, Y. (2010). Structure of the oligosaccharide region (core) of the lipopolysaccharides of Shigella flexneri types 2a and 5b. Russian Journal of Bioorganic Chemistry, 36(3), pp.396-399.

121. Kondo, A., Oketani, N., Maruyama, M., Taguchi, Y., Yamaguchi, Y., Miyao, H., Mashima, I., Oono, M., Wada, K., Tsuchiya, T., Takahashi, H. and Abe, S. (1998). Significance of serum surfactant protein-D (SP-D) level in patients with pulmonary tuberculosis. *Tuberculosis*, 73(10), pp.585-590.

122. Koopmans, J., van der Zee, J., Krop, E., Lopuhaa, C., Jansen, H. and Batenburg, J. (2004). Serum surfactant protein D is elevated in allergic patients. Clin Exp Allergy, 34(12), pp.1827-1833.

123. Korfhagen, T., Sheftelyevich, V., Burhans, M., Bruno, M., Ross, G., Wert, S., Stahlman, M., Jobe, A., Ikegami, M., Whitsett, J. and Fisher, J. (1998). Surfactant Protein-D Regulates Surfactant Phospholipid Homeostasis in Vivo. Journal of Biological Chemistry, 273(43), pp.28438-28443.

124. Kosma, P., Schulz, G. and Brade, H. (1988). Synthesis of a trisaccharide of 3-deoxy-d-manno-2-octulopyranosylonic acid (KDO) residues related to the genus-specific lipopolysaccharide epitope of Chlamydia. Carbohydrate Research, 183(2), pp.183-199.

125. Kostina, E., Ofek, I., Crouch, E., Friedman, R., Sirota, L., Klinger, G., Sahly, H. and Keisari, Y. (2005). Noncapsulated Klebsiella pneumoniae Bearing Mannose-Containing O Antigens Is Rapidly Eradicated from Mouse Lung and Triggers Cytokine Production by Macrophages following Opsonization with Surfactant Protein D. Infection and Immunity, 73(12), pp.8282-8290.

126. Kramer, R., Bella, J., Mayville, P. and Brodsky, B., Berman, H. (1999). Sequence dependent conformational variations of collagen triple-helical structure. Nat. Struct Biol., 6(5), pp.454-457.

127. Kramer, R., Vitagliano, L., Bella, J., Berisio, R., Mazzarella, L., Brodsky, B., Zagari, A. and Berman, H. (1998). X-ray crystallographic determination of a collagen-like peptide with the repeating sequence (Pro-Pro-Gly). *Journal of Molecular Biology*, 280(4), pp.623-638.

128. Kuan, S., Rust, K. and Crouch, E. (1992). Interactions of surfactant protein D with bacterial lipopolysaccharides. Surfactant protein D is an *Escherichia coli*-binding protein in bronchoalveolar lavage. *Journal of Clinical Investigation*, 90(1), pp.97-106.

129. Kucejko, W., Chyczewska, E., Naumnik, W. and Ossolińska, M. (2009). Concentration of surfactant protein D, Clara cell protein CC-16 and IL-10 in bronchoalveolar lavage (BAL) in

patients with sarcoidosis, hypersensivity pneumonitis and idiopathic pulmonary fibrosis. Folia Histochem Cytobiol., 47(2), pp.225-230.

130. Kudo, K., Sano, H., Takahashi, H., Kuronuma, K., Yokota, S., Fujii, N., Shimada, K., Yano, I., Kumazawa, Y., Voelker, D., Abe, S. and Kuroki, Y. (2004). Pulmonary Collectins Enhance Phagocytosis of *Mycobacterium avium* through Increased Activity of Mannose Receptor. The *Journal of Immunology*, 172(12), pp.7592-7602.

131. Kuroki, Y. and Sano, H. (1999). Functional Roles and Structural Analysis of Lung Collectins SP-A and SP-D. *Biology of the Neonate*, 76(Suppl. 1), pp.19-21.

132. Kuroki, Y., Takahashi, H., Chiba, H. and Akino, T. (1998). Surfactant proteins A and D: disease markers. Biochimica et Biophysica Acta - Molecular Basis of Disease, 1408(2-3), pp.334-345.

133. LaForce, F., Kelley, W. and Huber, G. (1973) Inactivation of staphylococci by alveolar macrophages with preliminary observations on the importance of alveolar lining material. Am. Rev. Respir. Dis., 108(4), pp.784-790.

134. Lahti, M., Löfgren, J., Marttila, R., Renko, M., Klaavuniemi, T., Haataja, R., Rämet, M. and Hallman, M. (2002). Surfactant Protein D Gene Polymorphism Associated with Severe Respiratory Syncytial Virus Infection. Pediatr Res, 51(6), pp.696-699.

135. Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., Valentin, F., Wallace, I., Wilm, A., Lopez, R., Thompson, J., Gibson, T. and Higgins, D. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, 23(21), pp.2947-2948.

136. Lee, Y., Lee, R., Rice, K., Ichikawa, Y. and Wong, T. (1991). Topography of binding sites of animal lectins: ligands' view. *Pure and Applied Chemistry*, 63(4), pp.499-506.

137. Lekkala, M., LeVine, A., Linke, M., Crouch, E., Linders, B., Brummer, E. and Stevens, D. (2006). Effect of Lung Surfactant Collectins on Bronchoalveolar Macrophage Interaction with *Blastomyces dermatitidis*: Inhibition of Tumor Necrosis Factor Alpha Production by Surfactant Protein D. *Infection and Immunity*, 74(8), pp.4549-4556.

138. Leslie, A. G. W., and Powel, H. R. (2007). Processing Diffraction Data with Mosflm. Evolving Methods for Macromolecular Crystallography, 254, pp.41-51.

139. Leth-Larsen, R., Garred, P., Jensenius, H., Meschi, J., Hartshorn, K., Madsen, J., Tornoe, I., Madsen, H., Sorensen, G., Crouch, E. and Holmskov, U. (2005). A Common Polymorphism in the SFTPD Gene Influences Assembly, Function, and Concentration of Surfactant Protein D. The Journal of Immunology, 174(3), pp.1532-1538.

140. Leth-Larsen, R., Homskov, U. and Højrup, P. (1999). Structural characterization of human and bovine lung surfactant protein D. *Biochemistry*, 343(3), pp.645-652.

141. Leth-Larsen, R., Nordenbaek, C., Tornoe, I., Moeller, V., Schlosser, A., Koch, C., Teisner, B., Junker, P. and Holmskov, U. (2003). Surfactant protein D (SP-D) serum levels in patients with community-acquired pneumonia. Clinical Immunology, 108(1), pp.29-37.

142. LeVine, A., Whitsett, J., Hartshorn, K., Crouch, E. and Korfhagen, T. (2001). Surfactant Protein D Enhances Clearance of Influenza A Virus from the Lung In Vivo. The Journal of Immunology, 167(10), pp.5868-5873.

143. Ley, K. (2003). The role of selectins in inflammation and disease. *Trends in Molecular Medicine*, 9(6), pp.263-268.

144. Lim, B., Wang, J., Holmskov, U., Hoppe, H. and Reid, K. (1994). Expression of the Carbohydrate Recognition Domain of Lung Surfactant Protein D and Demonstration of Its Binding to Lipopolysaccharides of Gram-Negative Bacteria. *Biochemical and Biophysical Research Communications*, 202(3), pp.1674-1680.

145. Lin, F., Chen, Y. and Chang, S. (2008). Clinical Importance of Bronchoalveolar Lavage Fluid and Blood Cytokines, Surfactant Protein D, and Kerbs von Lungren 6 Antigen in Idiopathic Pulmonary Alveolar Proteinosis. Mayo Clinic Proceedings, 83(12), pp.1344-1349.

146. Lu, J. (2002). Collectins and ficolins: sugar pattern recognition molecules of the mammalian innate immune system. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1572(2-3), pp.387-400.

147. Lunin, V. (1988). Use of the information on electron density distribution in macromolecules. *Acta Cryst Sect A*, 44(2), pp.144-150.

148. Mackay, A. and Pawley, G. (1963). Bravais lattices in four-dimensional space. *Acta Crystallographica*, 16(1), pp.11-19.

149. Madan, T., Kishore, U., Singh, M., Strong, P., Clark, H., Hussain, E., Reid, K. and Sarma, P. (2001). Surfactant proteins A and D protect mice against pulmonary hypersensitivity induced by *Aspergillus fumigatus* antigens and allergens. *Journal of Clinical Investigation*, 107(4), pp.467-475.

150. Madan,T., Eggleton, P., Kishore, U., S trong,P., Aggrawal, S., Sarma, P., and Reid, K.B. (1997). Binding of pulmonary surfactant proteins A and D to *Aspergillus fumigatus* conidia enhances phagocytosis and killing by human neutrophils and alveolar macrophages. *Infect. Immun*. 65, 3171–3179.

151. Madsen, J., Kliem, A., Tornoe, I., Skjodt, K., Koch, C. and Holmskov, U. (2000). Localization of Lung Surfactant Protein D on Mucosal Surfaces in Human Tissues. *The Journal of Immunology*, 164(11), pp.5866-5870.

152. Magalhães, P., Lopes, A., Mazzola, P., Rangel-Yagui, C., Penna, T. and Pessoa, A. (2007). Methods of endotoxin removal from biological preparations: a review. *J Pharm Pharm Sci*., 10(3), pp.388-404.

153. Mangoni, M., Epand, R., Rosenfeld, Y., Peleg, A., Barra, D., Epand, R. and Shai, Y. (2008). Lipopolysaccharide, a Key Molecule Involved in the Synergism between Temporins in Inhibiting Bacterial Growth and in Endotoxin Neutralization. Journal of Biological Chemistry, 283(34), pp.22907-22917.

154. Mansfield, L., Billett, E., Olsen, E. and Forsythe, S. (1996). Variation in Salmonella core lipopolysaccharide as detected by the monoclonal antibody M105. Letters in Applied Microbiology, 23(2), pp.104-106.

155. Manuel García-Ruiz, J. (2003). Nucleation of protein crystals. Journal of Structural Biology, 142(1), pp.22-31.

156. Mayo, D., Pike, R. and Trumper, P. (1994). Microscale organic laboratory. New York: Wiley.

157. McCormack, F. and Whitsett, J. (2002). The pulmonary collectins, SP-A and SP-D, orchestrate innate immunity in the lung. *Journal of Clinical Investigation*, 109(6), pp.707-712.

158. McCormack, F.X., Kuroki, Y., Stewart, J.J., Mason, R.J. and Voelker, D.R. (1994) Surfactant protein A amino acids Glu195 and Arg197 are essential for receptor binding, phospholipid aggregation, regulation of secretion, and the facilitated uptake of phospholipid by type II cells. *J. Biol. Chem*. 269(2), pp.9801–29807.

159. McPherson, A., Kuznetsov, Y., Malkin, A. and Plomp, M. (2003). Macromolecular crystal growth as revealed by atomic force microscopy. *Journal of Structural Biology*, 142(1), pp.32-46.

160. Medzhitov, R. and Janeway, C. (2000). Innate immune recognition: mechanisms and pathways. *Immunol Rev*, 173(1), pp.89-97.

161. Meschi, J. (2005). Surfactant protein D binds to human immunodeficiency virus (HIV) envelope protein gp120 and inhibits HIV replication. *Journal of General Virology*, 86(11), pp.3097-3107.

162. Mikhail, I., Yildirim, H., Lindahl, E. and Schweda, E. (2005). Structural characterization of lipid A from nontypeable and type f Haemophilus influenzae: Variability of fatty acid substitution. Analytical Biochemistry, 340(2), pp.303-316.

163. Miyamura, K., Malhotra, R., Hoppe, H., Reid, K., Phizackerley, P., Macpherson, P. and Bernal, A. (1994). Surfactant proteins A (SP-A) and D (SP-D): Levels in human amniotic fluid and localization in the fetal membranes. Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism, 1210(3), pp.303-307.

164. Moré, J., Voelker, D., Silveira, L., Edwards, M., Chan, E. and Bowler, R. (2010). Smoking reduces surfactant protein D and phospholipids in patients with and without chronic obstructive pulmonary disease. BMC Pulm Med, 10(1), p.53.

165. Mori, S. and Barth, H. (1999). Size exclusion chromatography. Berlin: Springer.

166. Morrison, D. and Leive, L. (1975). Fractions of lipopolysaccharide from Escherichia coli O111:B4 prepared by two extraction procedures. Journal of Biological Chemistry, 250(8), pp.2911-2919.

167. Murshudov, G., Vagin, A. and Dodson, E. (1997). Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallogr D Biol Cryst Struct Commun*, 53(3), pp.240-255.

168. Nadesalingam, J., Bernal, A., Dodds, A., Willis, A., Mahoney, D., Day, A., Reid, K. and Palaniyar, N. (2003). Identification and Characterization of a Novel Interaction between Pulmonary Surfactant Protein D and Decorin. Journal of Biological Chemistry, 278(28), pp.25678-25687.

169. Nayak, A., Dodagatta-Marri, E., Tsolaki, A. and Kishore, U. (2012). An Insight into the Diverse Roles of Surfactant Proteins, SP-A and SP-D in Innate and Adaptive Immunity. Frontiers in Immunology, 3.

170. Nigou, J., Gilleron, M. and Puzo, G. (2003). Lipoarabinomannans: from structure to biosynthesis. Biochimie, 85(1-2), pp.153-166.

171. Nikolaidis, N., White, M., Allen, K., Tripathi, S., Qi, L., McDonald, B., Taubenberger, J., Seaton, B., McCormack, F., Crouch, E. and Hartshorn, K. (2014). Mutations flanking the carbohydrate binding site of surfactant protein D confer antiviral activity for pandemic influenza A viruses. AJP: Lung Cellular and Molecular Physiology, 306(11), pp.L1036-L1044.

172. Oberley, R., Goss, K., Ault, K., Crouch, E. and Snyder, J. (2004). Surfactant protein D is present in the human female reproductive tract and inhibits Chlamydia trachomatis infection. Molecular Human Reproduction, 10(12), pp.861-870.

173. Ohtani, K., Suzuki, Y., Eda, S., Kawai, T., Kase, T., Yamazaki, H., Shimada, T., Keshi, H., Sakai, Y., Fukuoh, A., Sakamoto, T. and Wakamiya, N. (1999). Molecular Cloning of a Novel Human Collectin from Liver (CL-L1). *Journal of Biological Chemistry*, 274(19), pp.13681-13689.

174. Okita, J., MacDonald, P. and Jonston, J. (1982) Mobilization of arachidonic acid from specific glycerophospholipids of human fetal membranes during early labor. *Journal of Biological Chemistry,* 257(23), pp.14029-14034.

175. Olofsson, S., Kumlin, U., Dimock, K. and Arnberg, N. (2005). Avian influenza and sialic acid receptors: more than meets the eye? The Lancet Infectious Diseases, 5(3), pp.184-188.

176. Ooi, E., Wormald, P., Carney, A., James, C. and Tan, L. (2007). Surfactant Protein D Expression in Chronic Rhinosinusitis Patients and Immune Responses In Vitro to Aspergillus and Alternaria in a Nasal Explant Model. The Laryngoscope, 117(1), pp.51-57.

177. Orgeig, S., Hiemstra, P., Veldhuizen, E., Casals, C., Clark, H., Haczku, A., Knudsen, L. and Possmayer, F. (2010). Recent advances in alveolar biology: Evolution and function of alveolar proteins. Respiratory Physiology & Neurobiology, 173, pp.S43-S54.

178. Ortega, X., Silipo, A., Saldias, M., Bates, C., Molinaro, A. and Valvano, M. (2009). Biosynthesis and Structure of the Burkholderia cenocepacia K56-2 Lipopolysaccharide Core Oligosaccharide: Truncation of the core oligosaccharide leads to increased binding and sensitivity to Polymyxin B. Journal of Biological Chemistry, 284(32), pp.21738-21751.

179. Osiyemi, O. and Dickinson, G. (2000). Gram-positive pneumonia. Curr Infect Dis Rep, 2(3), pp.207-214.

180. Ozinsky, A., Underhill, D., Fontenot, J., Hajjar, A., Smith, K., Wilson, C., Schroeder, L. and Aderem, A. (2000). The repertoire for pattern recognition of pathogens by the innate immune

system is defined by cooperation between Toll-like receptors. *Proceedings of the National Academy of Sciences*, 97(25), pp.13766-13771.

181. Paananen, R., Glumoff, V., Sormunen, R. and Hallman, M. (1999). Genes of Surfactant Proteins SP-A, SP-B and SP-D are expressed in Eustachian Tube. Pediatr Res, 45(4, Part 2 of 2), pp.354A-354A.

182. Palaniyar, N., Clark, H., Nadesalingam, J, Hawgood, S. and Reid, K. (2003). Surfactant Protein D Binds Genomic DNA and Apoptotic Cells, and Enhances Their Clearance, in Vivo. Annals of the New York Academy of Sciences, 1010(1), pp.471-475.

183. Palaniyar, N., Nadesalingam, J., Clark, H., Shih, M., Dodds, A. and Reid, K. (2004). Nucleic Acid Is a Novel Ligand for Innate, Immune Pattern Recognition Collectins Surfactant Proteins A and D and Mannose-binding Lectin. *Journal of Biological Chemistry*, 279(31), pp.32728-32736.

184. Park, B., Song, D., Kim, H., Choi, B., Lee, H. and Lee, J. (2009). The structural basis of lipopolysaccharide recognition by the TLR4–MD-2 complex. Nature, 458(7242), pp.1191-1195.

185. Percy, M. and Gründling, A. (2014). Lipoteichoic Acid Synthesis and Function in Gram-Positive Bacteria. Annu. Rev. Microbiol., 68(1), pp.81-100.

186. Phillips, N., Apicella, M., Griffiss, J. and Gibson, B. (1992). Structural characterization of the cell surface lipooligosaccharides from a nontypable strain of Haemophilus influenzae. Biochemistry, 31(18), pp.4515-4526.

187. Phillips, A. and Signs, M. (2001). Desalting, Concentration, and Buffer Exchange by Dialysis and Ultrafiltration. Current Protocols in Protein Science.

188. Potterton, E., Briggs, P., Turkenburg, M. and Dodson, E. (2003). A graphical user interface to the CCP4 program suite. *Acta Cryst Sect D*, 59(7), pp.1131-1137.

189. Q. Ashton Acton, P. (2012). Coronaviridae Infections-Advances in Research and Treatment: 2012 Edition. ScholarlyEditions.

190. Qaseem, A., Sonar, S., Mahajan, L., Madan, T., Sorensen, G., Shamji, M. and Kishore, U. (2013). Linking surfactant protein SP-D and IL-13: Implications in asthma and allergy. Molecular Immunology, 54(1), pp.98-107.

191. Qiao, S., Luo, Q., Zhao, Y., Zhang, X. and Huang, Y. (2014). Structural basis for lipopolysaccharide insertion in the bacterial outer membrane. Nature, 511(7507), pp.108-111.

192. Randall, T. (2010). Bronchus-Associated Lymphoid Tissue (BALT). Mucosal Immunity, pp.187-241.

193. rcsb.org. Anon, (2015). [online] Available at: RSCB Bank (2015). RCSB PDB - Holdings Report. [online] Available at: http://www.rcsb.org/pdb/statistics/holdings.do [Accessed 5 Apr. 2015].

194. Read, R. J. (1990). Structure-factor probabilities for related structures. *Acta Crystallographica*., 46, pp.900–912.

195. Reading, P., Holmskov, U., and Anders, E (1998). Antiviral activity of bovine collectins against rotaviruses. *J. Gen.Virol*. 79(9), pp.2255–2263.

196. Restrepo, C., Dong, Q., Savov, J., Mariencheck, W. and Wright, J. (1999). Surfactant Protein D Stimulates Phagocytosis of *Pseudomonas aeruginosa* by Alveolar Macrophages. *Am. J. Respir. Cell Mol. Biol.*, 21(5), pp.576-585.

197. Řezáèová, P. (2002). Crystallography of Biological Macromolecules. *Materials Structure*, 9(1), pp.15-17

198. Rhodes, G. (2006). *Crystallography made crystal clear*. Amsterdam: Elsevier mgAcademic Press.

199. Ricciardolo, F., Sterk, P., Gaston, B. and Folkerts, G. (2004). Nitric oxide in health and disease of the respiratory system. *Physiol Rev.,* 84(3), pp.731–765

200. Richardson, M. (1993). Current Topics in Medical Mycology, volume 4: Edited by M. Borgers, R. Hay and M. G. Rinaldi. ISBN 3-540-97504-7. Springer-Verlag, New York.

201. Rixon, H., Brown, C., Brown, G. and Sugrue, R. (2002). Multiple glycosylated forms of the respiratory syncytial virus fusion protein are expressed in virus-infected cells. *J Gen Virol.,* 83(1), pp.61-66.

202. Rupp, B. (2010). Biomolecular Crystallography: Principals, Practice and Application to Structural Biology. New York, NY: Garland Science.

203. Sahly, H., Ofek, I., Podschun, R., Brade, H., He, Y., Ullmann, U. and Crouch, E. (2002). Surfactant Protein D Binds Selectively to Klebsiella pneumoniae Lipopolysaccharides Containing Mannose-Rich O-Antigens. The Journal of Immunology, 169(6), pp.3267-3274.

204. Samuel, G. and Reeves, P. (2003). Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. Carbohydrate Research, 338(23), pp.2503-2519.

205. Sasaki, R., Soejima, T., Matsumoto, A., Maruta, T., Yamada, K., Ota, Y., Kawabe, T., Nishimura, H., Sakai, E., Ejima, Y. and Sugimura, K. (2001). Clinical significance of serum pulmonary surfactant proteins A and D for the early detection of radiation pneumonitis. International Journal of Radiation Oncology, 50(2), pp.301-307.

206. Sawada, K., Ariki, S., Kojima, T., Saito, A., Yamazoe, M., Nishitani, C., Shimizu, T., Takahashi, M., Mitsuzawa, H., Yokota, S., Sawada, N., Fujii, N., Takahashi, H. and Kuroki, Y. (2010). Pulmonary Collectins Protect Macrophages against Pore-forming Activity of *Legionella pneumophila* and Suppress Its Intracellular Growth. *Journal of Biological Chemistry*, 285(11), pp.8434-8443.

207. Schagat, T., Wofford, J. and Wright, J. (2001). Surfactant Protein A Enhances Alveolar Macrophage Phagocytosis of Apoptotic Neutrophils. *The Journal of Immunology*, 166(4), pp.2727-2733.

208. Schelenz, S., Malhotra, R., Sim, R., Holmskov, U. and Bancroft, G. (1995). Binding of host collectins to the pathogenic yeast Cryptococcus neoformans: human surfactant protein D acts as an agglutinin for acapsular yeast cells. *Infect.Immun*. 63(9), pp3360–3366.

209. Schramek, S., Radziejewska-Lebrecht, J. and Mayer, H. (1985). 3-C-Branched aldoses in lipopolysaccharide of phase I Coxiella buurnetii and their role as immunodominant factors. Eur J Biochem, 148(3), pp.455-461.

210. Schroder, K. (2003). Interferon-gamma: an overview of signals, mechanisms and functions. Journal of Leukocyte Biology, 75(2), pp.163-189.

211. Seaton, B., Crouch, E., McCormack, F., Head, J., Hartshorn, K. and Mendelsohn, R. (2010). Review: Structural determinants of pattern recognition by lung collectins. *Innate Immunity*, 16(3), pp.143-150.

212. Sheldrick, G. (2010). Experimental phasing with SHELXC mg D / E : combining chain tracing with density modification. *Acta Cryst Sect D*, 66(4), pp.479-485.

213. Shepherd, V. (2002). Distinct Roles for Lung Collectins in Pulmonary Host Defense*. Am. J. Respir. Cell Mol. Biol.*, 26(3), pp.257-260.

214. Shockman, G. and Barren, J. (1983). Structure, Function, and Assembly of Cell Walls of Gram-Positive Bacteria. Annu. Rev. Microbiol., 37(1), pp.501-527.

215. Shrive, A., Martin, C., Burns, I., Paterson, J., Martin, J., Townsend, J., Waters, P., Clark, H., Kishore, U., Reid, K. and Greenhough, T. (2009). Structural Characterisation of Ligand-Binding Determinants in Human Lung Surfactant Protein D: Influence of Asp325. *Journal of Molecular Biology*, 394(4), pp.776-788.

216. Shrive, A., Tharia, H., Strong, P., Kishore, U., Burns, I., Rizkallah, P., Reid, K. and Greenhough, T. (2003). High-resolution Structural Insights into Ligand binding and Immune Cell Recognition by Human Lung Surfactant Protein D. *Journal of Molecular Biology*, 331(2), pp.509-523.

217. Sigma-Aldrich, (2015). N-Linked Glycans. [online] Available at: http://www.sigmaaldrich.com/technical-documents/articles/biology/glycobiology/n-glycans.html [Accessed 14 Apr. 2015].

218. Sin, D., Man, S., McWilliams, A. and Lam, S. (2008). Surfactant Protein D and Bronchial Dysplasia in Smokers at High Risk of Lung Cancer. Chest, 134(3), pp.582-588.

219. Sioud, S., Jahouh, F., Nashed, M., Joly, N. and Banoub, J. (2010). Determination of distinctive carbohydrate signatures obtained from the Aeromonas hydrophila (chemotype II) core oligosaccharide pinpointing the presence of the 4-O-phosphorylated 5-O-linked Kdo reducing end group using electrospray ionization quadrupole orthogonal time-of-flight mass spectrometry and tandem mass spectrometry. Rapid Commun. Mass Spectrom., 24(17), pp.2475-2490.

220. Sorensen, G., Madsen, J., Kejling, K., Tornoe, I., Nielsen, O., Townsend, P., Poulain, F., Nielsen, C., Reid, K., Hawgood, S., Falk, E. and Holmskov, U. (2006). Surfactant protein D is proatherogenic in mice. AJP: Heart and Circulatory Physiology, 290(6), pp.2286-2294.

221. Stamler, J., Lamas, S. and Fang, F. (2001). Nitrosylation: the prototypic redox based signalling mechanism. Cell, 106(6), pp.675-683.

222. Stenutz, R., Weintraub, A. and Widmalm, G. (200co6). The structures of Escherichia coli O-polysaccharide antigens. FEMS Microbiology Reviews, 30(3), pp.382-403.

223. Strong, P., Reid, K. and Clark, H. (2002). Intranasal delivery of a truncated recombinant human SP-D is effective at down-regulating allergic hypersensitivity in mice sensitized to allergens of *Aspergillus fumigatus*. *Clinical and Experimental Immunology*, 130(1), pp.19-24.

224. Strong, P., Townsend, P., Mackay, R., Reid, K. and Clark, H. (2003). A recombinant fragment of human SP-D reduces allergic responses in mice sensitized to house dust mite allergens. Clinical and Experimental Immunology, 134(2), pp.181-187.

225. Stuart, G., Lynch, N., Day, A., Schwaeble, W. and Sim, R. (1997). The C1q and collectin binding site within C1 q receptor (cell surface calreticulin). *Immunopharmacology*, 38(1-2), pp.73-80.

226. Sutcliffe, I. and Shaw, N. (1991). Atypical lipoteichoic acids of gram-positive bacteria. *J Bacteriol*., 173(22), pp.7065-7069.

227. Takahashi, H., Imai, Y., Fujishima, T., Shiratori, M., Murakami, S., Chiba, H., Kon, H., Kuroki, Y. and Abe, S. (2001). Diagnostic significance of surfactant proteins A and D in sera from patients with radiation pneumonitis. European Respiratory Journal, 17(3), pp.481-487.

228. Terwilliger, T. and Berendzen, J. (1996). Bayesian Weighting for Macromolecular Crystallographic Refinement. *Acta Crystallogr D Biol Cryst Struct Commun*, 52(4), pp.743-748.

229. Tino, M. and Wright, J. (1998). Interactions of surfactant protein A with epithelial cells and phagocytes. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1408(2-3), pp.241-263.

230. Trifiro, S., Bourgault, A., Lebel, F. and René, P. (1990). Ghost *Mycobacteria* on Gram Stain. *Journal of Clinical Microbiology*, 28(1), pp.146-147.

231. Tucker, J. (1999). Historical Trends Related to Bioterrorism: An Empirical Analysis. Emerg. Infect. Dis., 5(4), pp.498-504.

232. UniProt: a hub for protein information. (2014). *Nucleic Acids Research*, 43(D1), pp.D204-D212.

233. Urh, M., Simpson, D. and Zhao, K. (2009). Chapter 26 Affinity Chromatography. Guide to Protein Purification, 2nd Edition, pp.417-438.

234. Van De Wetering, J., Van Golde, L. and Batenburg, J. (2004). Collectins. *European Journal of Biochemistry*, 271(7), pp.1229-1249.

235. van de Wetering, J., van Remoortere, A., Vaandrager, A., Batenburg, J., van Golde, L., Hokke, C. and van Hellemond, J. (2004b). Surfactant Protein D Binding to Terminal α1-3–Linked Fucose Residues and to Schistosoma mansoni. Am J Respir Cell Mol Biol, 31(5), pp.565-572.

236. Van de Wetering, J., Van Eijk, M., Van Golde, L., Hartung, T., van Strijp, J. and Batenburg, J. (2001). Characteristics of Surfactant Protein A and D Binding to Lipoteichoic Acid and

Peptidoglycan, 2 Major Cell Wall Components of Gram-Positive Bacteria. *The Journal of Infectious Diseases*, 184(9), pp.1143-1151.

237. van Rozendaal, B., van Spriel, A., van de Winkel, J. and Haagsman, H. (2000). Role of Pulmonary Surfactant Protein D in Innate Defense against *Candida albicans*. *The Journal of Infectious Diseases*, 182(3), pp.917-922.

238. Varga, J., Denton, C. and Wigley, F. (2012). Scleroderma. New York: Springer.

239. Varki, A. (1999). *Essentials of glycobiology*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

240. Veldhuizen, E., van Eijk, M. and Haagsman, H. (2011). The carbohydrate recognition domain of collectins. *FEBS Journal*, 278(20), pp.3930-3941.

241. Vigerust, D., Ulett, K., Boyd, K., Madsen, J., Hawgood, S. and McCullers, J. (2007). N-Linked Glycosylation Attenuates H3N2 Influenza Viruses. Journal of Virology, 81(16), pp.8593-8600.

242. Vinogradov, E., Lindner, B., Kocharova, N., Senchenkova, S., Shashkov, A., Knirel, Y., Holst, O., Gremyakova, T., Shaikhutdinova, R. and Anisimov, A. (2002). The core structure of the lipopolysaccharide from the causative agent of plague, Yersinia pestis. Carbohydrate Research, 337(9), pp.775-777.

243. Vinogradov, E., Frirdich, E., MacLean, L., Perry, M., Peterson, B., Duus, J. and Whitfield, C. (2002b) Structures of Lipopolysaccharides from *Klebsiella Pneumoniae*. Elucidation of the Structure of the Linkage Region between Core and Polysaccharide O-chain and Identification of the Residues at the Non-Reducing Termini of the O-chain. *Journal of Biological Chemistry* 277(28), pp.25070-25081.

244. Vinogradov, E., Petersen, B., Duus, J. and Radziejewska-Lebrecht, J. (2003). The structure of the polysaccharide part of the LPS from Serratia marcescens serotype O19, including linkage region to the core and the residue at the non-reducing end. Carbohydrate Research, 338(23), pp.2757-2761.

245. Wallis, R. and Drickamer, K. (1997). Asymmetry adjacent to the collagen-like domain in rat liver mannose-binding protein. *Biochemistry*, 325, pp.391-400.

246. Wang, G. and Lu, T. (2013). Crystal Lattices and Reciprocal Lattices. *RHEED Transmission Mode and Pole Figures*, pp.7-22.

247. Wang, H., Head, J., Kosma, P., Brade, H., Müller-Loennies, S., Sheikh, S., McDonald, B., Smith, K., Cafarella, T., Seaton, B. and Crouch, E. (2008). Recognition of Heptoses and the Inner Core of Bacterial Lipopolysaccharides by Surfactant Protein D. *Biochemistry*, 47(2), pp.710-720.

248. Wang, J., Kishore, U., Lim, B., Strong, P. and Reid, K. (1996). Interaction of human lung surfactant proteins A and D with mite (Dermatophagoides pteronyssinus) allergens. Clinical and Experimental Immunology, 106(2), pp.367-373.

249. Watts, D., Müller-Dieckmann, J., Tsakanova, G., Lamzin, V. and Groves, M. (2010). Quantitive evaluation of macromolecular crystallization experiments using 1,8-ANS fluorescence. *Acta Cryst Sect D*, 66(8), pp.901-908.

250. Weaver,T., and Whitsett, J. (1991). Function and regulation of expression of pulmonary surfactant-associated proteins. *Biochemistry* 273, pp.249–264.

251. Weik, M., Ravelli, R., Kryger, G., McSweeney, S., Raves, M., Harel, M., Gros, P., Silman, I., Kroon, J. and Sussman, J. (2000). Specific chemical and structural damage to proteins produced by synchrotron radiation. Proceedings of the National Academy of Sciences, 97(2), pp.623-628.

252. Weis, W. and Drickamer, K. (1994). Trimeric structure of a C-type mannose-binding protein. *Structure*, 2(12), pp.1227-1240.

253. Weis, W., Taylor, M. and Drickamer, K. (1998). The C-type lectin superfamily in the immune system. *Immunol Rev*, 163(1), pp.19-34.

254. Werling, D. and Jungi, T. (2003). TOLL-like receptors linking innate and adaptive immune response. Veterinary Immunology and Immunopathology, 91(1), pp.1-12.

255. Wertz, G. and Lichtenstein, D. (1988). Synthesis, processing and maturation of the respiratory syncytial virus surface glycoprotein, G. Virus Research, 11, p.4.

256. Weyer, C., Sabat, R., Wissel, H., Krüger, D., Stevens, P. and Prösch, S. (2000). Surfactant Protein A Binding to Cytomegalovirus Proteins Enhances Virus Entry into Rat Lung Cells. Am. J. Respir. Cell Mol. Biol., 23(1), pp.71-78.

257. Who.int, (2015). WHO | Schistosomiasis. [online] Available at: http://www.who.int/mediacentre/factsheets/fs115/en/ [Accessed 15 Apr. 2015].

258. Winkler, C., Atochina-Vasserman, E., Holz, O., Beers, M., Erpenbeck, V., Krug, N., Roepcke, S., Lauer, G., Elmlinger, M. and Hohlfeld, J. (2011). Comprehensive characterisation of pulmonary and serum surfactant protein D in COPD. Respiratory Research, 12(1), p.29.

259. Woodworth, B., Lathers, D., Neal, J., Skinner, M., Richardson, M., Young, M. and Schlosser, R. (2006). Immunolocalization of surfactant protein A and D in sinonasal mucosa. American Journal of Rhinology, 20(4), pp.461-465.

260. Wright, J. (2005). Immunoregulatory functions of surfactant proteins. Nature Reviews Immunology, 5(1), pp.58-68.

261. Wu, H., Kuzmenko, A., Wan, S., Schaffer, L., Weiss, A., Fisher, J., Kim, K. and McCormack, F. (2003). Surfactant proteins A and D inhibit the growth of Gram-negative bacteria by increasing membrane permeability. *Journal of Clinical Investigation*, 111(10), pp.1589-1602.

262. Wu, Y., Liu, Z., Wei, R., Pan, S., Mao, N., Chen, B., Han, J., Zhang, F., Holmskov, U., Xia, Z., de Groot, P., Reid, K., Xu, W. and Sorensen, G. (2009). Elevated Plasma Surfactant Protein D (SP-D) Levels and a Direct Correlation with Anti-severe Acute Respiratory Syndrome Coronavirus-specific IgG Antibody in SARS Patients. Scandinavian Journal of Immunology, 69(6), pp.508-515.

263. Yamazoe, M., Nishitani, C., Takahashi, M., Katoh, T., Ariki, S., Shimizu, T., Mitsuzawa, H., Sawada, K., Voelker, D., Takahashi, H. and Kuroki, Y. (2008). Pulmonary Surfactant Protein D

Inhibits Lipopolysaccharide (LPS)-induced Inflammatory Cell Responses by Altering LPS Binding to Its Receptors. Journal of Biological Chemistry, 283(51), pp.35878-35888.

264. Yeh, J., Du, S., Tortajada, A., Paulo, J. and Zhang, S. (2005). Peptergents: Peptide Detergents That Improve Stability and Functionality of a Membrane Protein, Glycerol-3-phosphate Dehydrogenase. *Biochemistry*, 44(51), pp.16912-16919.

265. Yethon, J., Vinogradov, E., Perry, M. and Whitfield, C. (2000). Mutation of the Lipopolysaccharide Core Glycosyltransferase Encoded by waaG destabilizes the Outer Membrane of Escherichia coli by Interfering with Core Phosphorylation. Journal of Bacteriology, 182(19), pp.5620-5623.

266. Yong, S., Vuk-Pavlovic, Z., Standing, J., Crouch, E. and Limper, A. (2003). Surfactant Protein D-Mediated Aggregation of *Pneumocystis carinii* Impairs Phagocytosis by Alveolar Macrophages. Infection and Immunity, 71(4), pp.1662-1671.

267. Zelensky, A. and Gy, J. (2005). The C-type lectin-like domain superfamily. *FEBS Journal*, 272(24), pp.6179-6217.

268. Zhang, P., McAlinden, A., Li, S., Schumacher, T., Wang, H., Hu, S., Sandell, L. and Crouch, E. (2001). The Amino-terminal Heptad Repeats of the Coiled-coil Neck Domain of Pulmonary Surfactant Protein D Are Necessary for the Assembly of Trimeric Subunits and Dodecamers. *Journal of Biological Chemistry*, 276(23), pp.19862-19870.

269. Zhou, Y., Lu, K., Pfefferle, S., Bertram, S., Glowacka, I., Drosten, C., Pohlmann, S. and Simmons, G. (2010). A Single Asparagine-Linked Glycosylation Site of the Severe Acute Respiratory Syndrome Coronavirus Spike Glycoprotein Facilitates Inhibition by Mannose-Binding Lectin through Multiple Mechanisms. Journal of Virology, 84(17), pp.8753-8764.

Appendix

Protein preparations used:

1. Affinity purified, 0.5ml, 11.29mg/ml (in TBS and 5mM EDTA)
2. Affinity purified + size exclusion chromatography, 0.45ml, 10.97mg/ml (in TBS and 5mM EDTA)
3. Affinity purified + size exclusion chromatography + endotoxin treatment, 0.6ml, 8.76mg/ml (in PBS)

Unless otherwise stated all protein to reservoir ratios in drops are 1µl:1µl.

Table 1

CCS1, 16/12/10

Non-variable conditions: 16% w/v PEG, 0.1M Tris

Variable conditions: molecular weight PEG, pH Tris

Protein prep 1 dialysed in 10mM Tris, 10mM $CaCl_2$ and 140mM NaCl at pH 7.5, **7.49mg/ml**

		1	2	3	4	5
	pH/PEG	2000	4000	6000	8000	10000
A	6					
B	7					
C	8					
D	8.5					

Table 2

CCS2, 17/12/10

Non-variable conditions: 16% w/v PEG, 0.1M Tris

Variable conditions: molecular weight PEG, pH Tris

Protein prep 3 **8.76mg/ml**, 0.5µl 50mM $CaCl_2$ added to each drop 1/2/11

Column 6 added 22/2/11. Calcium added to A6 and C6 1/3/11. Calcium added to B6 and D6 8/3/11

		1	2	3	4	5	6
	pH/PEG	2000	4000	6000	8000	10000	
A	6						6000 pH6
B	7						6000 pH6
C	8						10000 pH7
D	8.5						10000 pH7

Table 3

CCS3, 1/2/11

Non-variable conditions: 16% w/v PEG, 0.1M Tris

Variable conditions: molecular weight PEG, pH Tris

Protein prep 3 dialysed in 10mM Tris, 10mM $CaCl_2$ and 140mM NaCl at pH 7.5, **8.39mg/ml**

		1	2	3	4	5
	pH/PEG	2000	4000	6000	8000	10000
A	6					
B	7					
C	8					
D	8.5					

Table 4

CCS4, 23/2/11

Non-variable conditions: 16% w/v PEG, 0.1M Tris

Variable conditions: molecular weight PEG, pH Tris

Protein prep 2 dialysed in 10mM Tris, 10mM $CaCl_2$ and 140mM NaCl at pH 7.5, **8.16mg/ml**

		1	2	3	4	5
	pH/PEG	2000	4000	6000	8000	10000
A	6					
B	7					
C	8					
D	8.5					

Table 5

CCS5, 13/5/11

Non-variable conditions: 16% w/v PEG, 0.1M Tris

Variable conditions: molecular weight PEG, pH Tris

Protein prep 3 **8.76mg/ml**, 0.5µl 50mM $CaCl_2$ added to each drop 23/5/11

		1	2	3	4	5
	pH/PEG	2000	4000	6000	8000	10000
A	6					
B	7					
C	8					

Table 6

CCS6, 10/6/11

Non-variable conditions: 0.1M Tris

Variable conditions: % w/v PEG, molecular weight PEG, pH Tris

Wells A and B 1-3 Protein prep 1 dialysed in 10mM Tris, 10mM $CaCl_2$ and 140mM NaCl at pH 7.5, **7.49mg/ml**

Wells A and B 5-6 Protein prep 3 dialysed in 10mM Tris, 10mM $CaCl_2$ and 140mM NaCl at pH 7.5, **8.39mg/ml**

Wells C and D 5-6 Protein prep 2 dialysed in 10mM Tris, 10mM $CaCl_2$ and 140mM NaCl at pH 7.5, **8.16mg/ml**

	1	2	3	4	5	6
A	15% PEG 6K pH6	16% PEG 6K pH6	17% PEG 6K pH6		16% PEG 2K	16% PEG 4K
B	15% PEG 6K pH6	16% PEG 6K pH6	17% PEG 6K pH6		16% PEG 2K	16% PEG 4K
C					16% PEG 8K pH6	16% PEG 8K pH7
D					16% PEG 8K pH6	16% PEG 8K pH7

Table 7

CCS7, 7/7/11

Non-variable conditions: 16% w/v PEG, 0.1M Tris

Variable conditions: molecular weight PEG, pH Tris

Protein prep 1 dialysed in 10mM Tris, 10mM $CaCl_2$ and 140mM NaCl at pH 7.5, incubated with intact *Haemophilus influenzae* Eagan 4A lipopolysaccharide and reconcentrated

		1	2	3	4	5
	pH/PEG	2000	4000	6000	8000	10000
A	6					
B	7					
C	8					
D	8.5					

Table 8

CCS8, 5/12/11

Non-variable conditions: 16% w/v PEG, 0.1M Tris

Variable conditions: molecular weight PEG, pH Tris

Protein prep 3 **8.76mg/ml**, 0.5µl 50mM $CaCl_2$ added to each drop 13/12/11

		1	2	3	4	5
	pH/PEG	2000	4000	6000	8000	10000
A	6					
B	7					
C	8					
D	8.5					

Table 9

CCS9, 30/3/12

Non-variable conditions: 16% w/v PEG, 0.1M Tris

Variable conditions: molecular weight PEG, pH Tris

Protein prep 3 **8.76mg/ml**, 0.5µl 50mM $CaCl_2$ added to each drop 6/4/12

	1	2	3	4	5	6
A	16% PEG 6K pH6	16% PEG 8K pH7	16% PEG 10K pH6	16% PEG 4K pH8		
B	16% PEG 6K pH6	16% PEG 8K pH7	16% PEG 10K pH6	16% PEG 4K pH8		
C						
D						

Table 10

CCS10, 14/5/12

Non-variable conditions: 16% w/v PEG, 0.1M Tris

Variable conditions: molecular weight PEG, pH Tris

Protein prep 3 **8.76mg/ml**, 0.5µl 50mM $CaCl_2$ added to each drop 21/5/12

	1	2	3	4	5	6
A	16% PEG 6K pH6	16% PEG 8K pH7	16% PEG 10K pH6	16% PEG 4K pH8		
B	16% PEG 6K pH6	16% PEG 8K pH7	16% PEG 10K pH6	16% PEG 4K pH8		
C						
D						

Table 11

CCS11, 11/6/12

Non-variable conditions: 16% w/v PEG, 0.1M Tris

Variable conditions: molecular weight PEG, pH Tris

Protein prep 3 **8.76mg/ml**, 0.5µl 50mM $CaCl_2$ added to each drop 18/6/12

	1	2	3	4	5	6
A	16% PEG 6K pH6	16% PEG 8K pH7	16% PEG 10K pH6	16% PEG 4K pH8		
B	16% PEG 6K pH6	16% PEG 8K pH7	16% PEG 10K pH6	16% PEG 4K pH8		
C						
D						

Table 12

CCS12, 21/6/12

Non-variable conditions: 16% w/v PEG, 0.1M Tris

Variable conditions: molecular weight PEG, pH Tris

Protein prep 3 **8.76mg/ml**, 0.5µl 50mM $CaCl_2$ added to each drop as stated

Wells A and B 5-6, laid down 29/6/12, calcium added immediately

Wells C and D 5-6, laid down 15/8/12, calcium added 22/8/12

Wells C and D 2-3, laid down 27/9/12, calcium added immediately

	1	2	3	4	5	6
A	16% PEG 6K pH6	16% PEG 8K pH7	16% PEG 10K pH6	16% PEG 4K pH8	16% PEG 10K pH7	16% PEG 4K pH8
B	16% PEG 6K pH6	16% PEG 8K pH7	16% PEG 10K pH6	16% PEG 4K pH8	16% PEG 10K pH7	16% PEG 4K pH8
C		16% PEG 10K pH7	16% PEG 4K pH8		16% PEG 10K pH7	16% PEG 4K pH8
D		16% PEG 10K pH7	16% PEG 4K pH8		16% PEG 10K pH7	16% PEG 4K pH8

Table 13

CCS13, 11/7/12

Non-variable conditions: 16% w/v PEG, 0.1M Tris

Variable conditions: molecular weight PEG, pH Tris

Protein prep 3 **8.76mg/ml**, 0.5µl 50mM $CaCl_2$ added to each drop immediately

	1	2	3	4	5	6
A	16% PEG 10K pH7	16% PEG 10K pH7	16% PEG 4K pH8	16% PEG 4K pH8	16% PEG 10K pH7	16% PEG 4K pH8
B	16% PEG 10K pH7	16% PEG 10K pH7	16% PEG 4K pH8	16% PEG 4K pH8	16% PEG 10K pH7	16% PEG 4K pH8
C	16% PEG 10K pH7	16% PEG 4K pH8	16% PEG 10K pH7	16% PEG 4K pH8	16% PEG 10K pH7	16% PEG 4K pH8
D	16% PEG 10K pH7	16% PEG 4K pH8	16% PEG 10K pH7	16% PEG 4K pH8	16% PEG 10K pH7	16% PEG 4K pH8

Table 14

CCS14, 12/7/12

Non-variable conditions: 16% w/v PEG, 0.1M Tris

Variable conditions: molecular weight PEG, pH Tris

Protein prep 3 **6.52mg/ml**, 0.18mM hydrolysed *E. coli* B6 lipopolysaccharide and 11.1mM $CaCl_2$, incubated for 2 hours at room temperature.

		1	2	3	4	5
	pH/PEG	2000	4000	6000	8000	10000
A	6					
B	7					
C	8					
D	8.5					

Table 15

CCS15, 20/11/12

Non-variable conditions: 16% w/v PEG, 0.1M Tris

Variable conditions: molecular weight PEG, pH Tris

Protein prep 3 **8.76mg/ml**, 0.5µl 50mM $CaCl_2$ added to each drop immediately, excluding A and B 5-6 & C and D 1-4 where calcium was added 26/11/12; and C and D 3-4 after 24 hours.

	1	2	3	4	5	6
A	16% PEG 10K pH7	16% PEG 10K pH7	16% PEG 4K pH8	16% PEG 4K pH8	16% PEG 10K pH7	16% PEG 4K pH8
B	16% PEG 10K pH7	16% PEG 10K pH7	16% PEG 4K pH8	16% PEG 4K pH8	16% PEG 10K pH7	16% PEG 4K pH8
C	16% PEG 10K pH7	16% PEG 4K pH8	16% PEG 10K pH7	16% PEG 4K pH8	16% PEG 4K pH8	
D	16% PEG 10K pH7	16% PEG 4K pH8	16% PEG 10K pH7	16% PEG 4K pH8	16% PEG 4K pH8	

Table 16

CCS16, 7/2/12

Non-variable conditions: 16% w/v PEG, 0.1M Tris

Variable conditions: molecular weight PEG, pH Tris

Protein prep 3 **8.76mg/ml**, 0.5µl 50mM $CaCl_2$ added to each drops in column 1 after 1 day, column 2 after 4 days, columns 3 and 4 after 1 week, and to columns 5 and 6 immediately.

Wells C1/D1 protein prep 3 **7.42mg/ml** incubated with 0.18mM hydrolysed *S. minnesota* lipopolysaccharide and 10mM $CaCl_2$ 8/3/12

	1	2	3	4	5	6
A	16% PEG 10K pH7	16% PEG 4K pH8	16% PEG 10K pH7	16% PEG 4K pH8	16% PEG 10K pH7	16% PEG 4K pH8
B	16% PEG 10K pH7	16% PEG 4K pH8	16% PEG 10K pH7	16% PEG 4K pH8	16% PEG 10K pH7	16% PEG 4K pH8
C	16% PEG 10K pH7					
D	16% PEG 10K pH7					

Table 17 Details of crystals which were tested or soaked along with collected data sets and subsequent processing. Fast - automatically reduced by fast_dp and XIA2 immediately following data collection using the Automatic Software Pipeline; Full - data integrated and reduces through MOSFLM and scaled using SCALA. Resolution in Å.

Crystal	Procedure	Soak time (exchange to freeze)	Result		Processing (Fast/Full)
CCS2 B5 **TEST**	+ 1µl 6.7mM Eagan 4A LPS 28/4/11	-	-		
CCS2 A3 1 CCS2 A3 2	Exchanged and frozen Native 4/5/11	-	Snap 1.75 Å Snap 1.8 Å	7/5/11 I03	-
CCS2 A3 3 CCS2 A3 4 CCS2 A3 5	Exchanged and frozen + 12µl 6.7mM Eagan 4A LPS 4/5/11	- 1h18m 1h24m	Snap 9 Å Collected 1.8 Å Collected 1.8 Å	7/5/11 I03	- Fast 2.38 Fast 1.92
CCS2 A3 6 CCS2 A3 7 CCS2 A3 8	Reexchanged + 6µl 6.7mM Eagan 4A LPS 18/5/11	22m 27m 32m	Snap 2.5 Å Collected 1.5 Å Snap 2.86 Å	20/5/11 I04-1	- Fast 1.68 -
CCS2 B1 1 CCS2 B1 2 CCS2 B1 3 CCS2 B1 4	Exchanged and frozen 1/7/11	-	No diffraction	4/7/11 I04-1	
CCS8 B5 **TEST**	5-20% MPD buffer containing 30mM *E. coli* B4 PS 28/2/12	-	-		

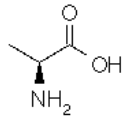
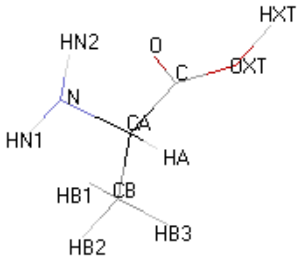
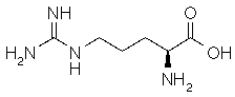
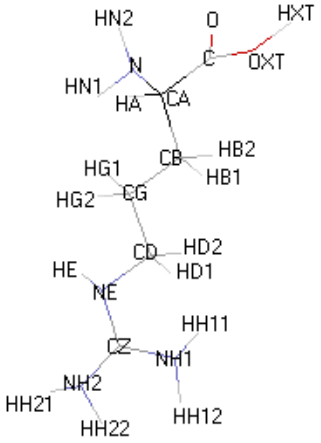
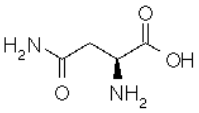
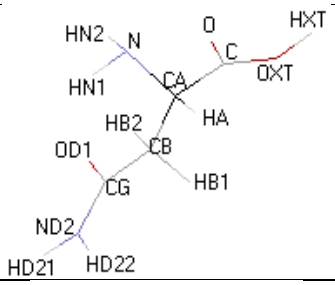
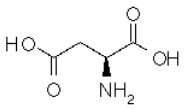
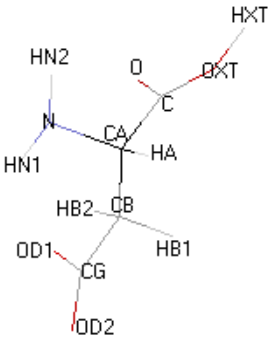
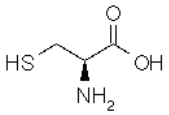
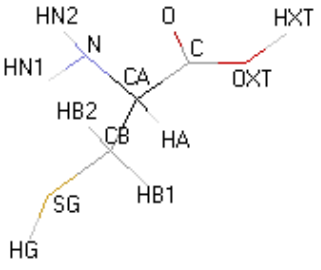
Table 18 Details of crystals which were tested or soaked along with collected data sets and subsequent processing. Fast - automatically reduced by fast_dp and XIA2 immediately following data collection using the Automatic Software Pipeline; Full - data integrated and reduces through MOSFLM and scaled using SCALA. Resolution in Å.

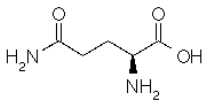
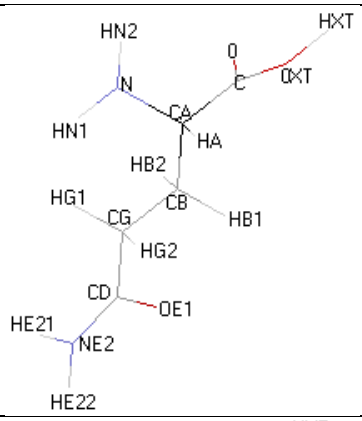
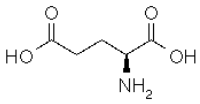
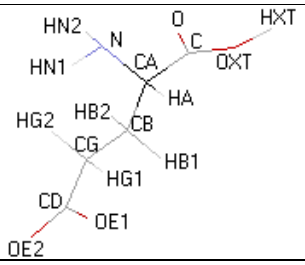
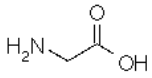
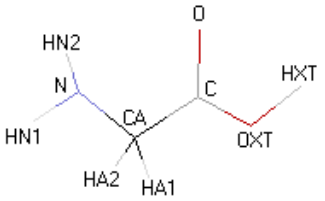
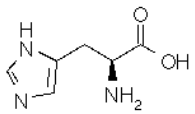
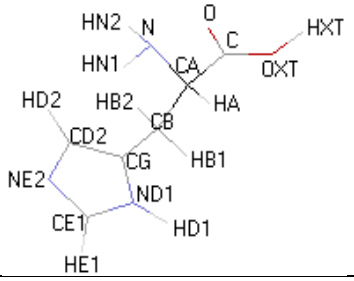
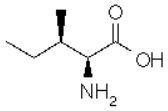
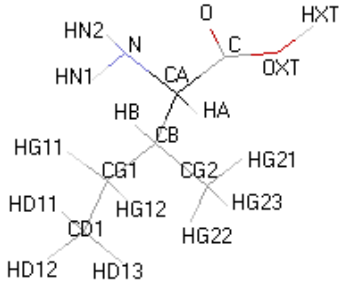
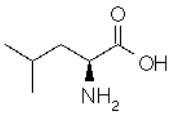
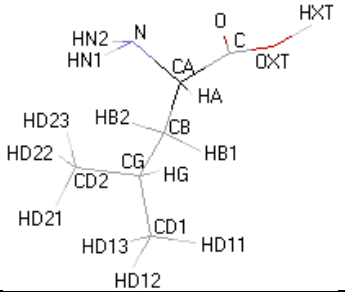
CCS8 C1 **TEST**	5-20% MPD buffer containing 30mM *E. coli* B4 PS 29/2/12	-	-		
CCS8 B4	5-20% MPD buffer containing 30mM *E. coli* B4 PS 6/3/12	-	Dissolved 7/3/12		
CCS2 A1 1 CCS2 A1 2	Moved to CCS8 C1	-	Dissolved instantly		
CCS3 C1 1 CCS3 C1 2	Exchanged (native) and frozen 7/3/12		DC 2.5 Å Snap 4.2 Å	6/5/12 I04	Unprocessed -
CCS4 C1	Moved to CCS8 C1	-	Dissolved after 4 hours		
CCS5 C1	Moved to CCS8 C1 7/3/12	-	Very degraded 15/3/12		
CCS8 B3	Moved to CCS8 C1	-	Dissolved instantly		
CCS2 C2 1 CCS2 C2 2 CCS2 C2 3 CCS2 C2 4 CCS2 C2 5	5-20% MPD buffer containing 30mM *E. coli* B4 PS Exchange after 3h 45m 15/3/12	5m 10m 13m 17m 23m	DC 2.4 Å DC 2 Å DC 2.4 Å DC 2.5 Å DC 2 Å	17/3/12 I03	Fast 3.14 Fast 2.79 Fast 3.05 Fast 3.55 Fast 2.98
CCS2 D6 1 CCS2 D6 2	5-20% MPD buffer containing 30mM *E. coli* B4 PS. Exchanged after 56m 15/3/12	9m 12m	DC 1.6 Å DC 1.7 Å	17/3/12 I03	Fast 1.74 Fast 2.39
CCS9 B3 1 CCS9 B3 2 CCS9 B3 3	5-20% MPD buffer containing 30mM *E. coli* B6 PS. Exchanged immediately. 4/5/12	36m 41m 44m	DC 1.9 DC 2.0 DC 1.9	6/5/12 I04	Full 2.09 Full 2.83 Full 2.68
CCS9 A4 1	5-20% MPD buffer containing 30mM *E. coli* B6 PS. Exchanged immediately. 4/5/12	10m	DC 2.0	6/5/12 I04	Unprocessed
CCS3 D4 1 CCS3 D4 1a CCS3 D4 2 CCS3 D4 3	5-20% MPD buffer containing 30mM *E. coli* J5 PS. Exchanged after 40 minutes. 17/7/12	44m 44m 46m 51m	DC 2 DC 2.5 DC 2.5 – 2.35 No diffraction	22/7/12 I04	Full 2.97 Fast 2.36 Full 3.45 -
CCS8 C 1	Re-exchanged with 20% MPD cryo containing 30mM *E. coli* J5 PS (10µl out, 10µl in) 17/7/12	1h5m	DC 1.5	22/7/12 I04	Full 1.65
CCS13 B6 1 CCS13 B6 2 CCS13 B6 3 CCS13 B6 4	5-20% MPD containing 30mM *E. coli* J5 PS. Exchanged after 40 minutes. 6/8/12	3m 5m 6m 9m	- DC 1.9 DC 2.14 DC 1.7	9/8/12 I03	- Unprocessed Unprocessed Unprocessed

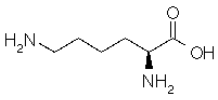
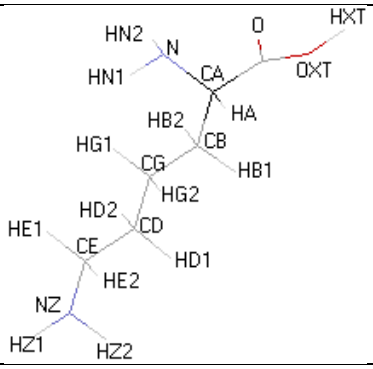
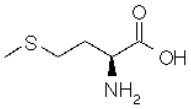
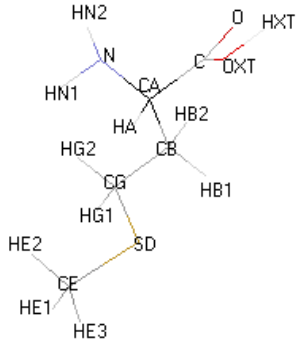
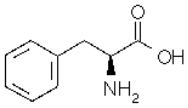
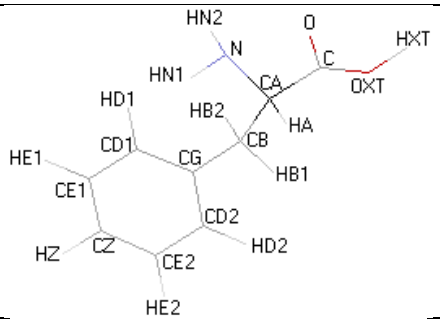
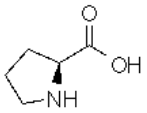
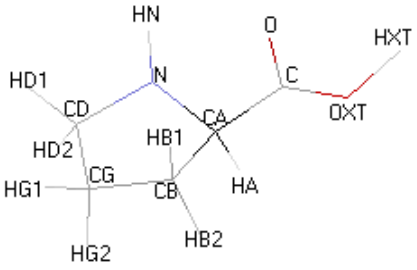
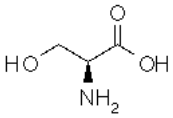
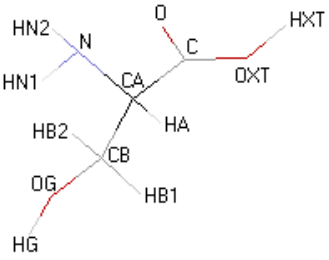
Table 19 Details of crystals which were tested or soaked along with collected data sets and subsequent processing. Fast - automatically reduced by fast_dp and XIA2 immediately following data collection using the Automatic Software Pipeline; Full - data integrated and reduces through MOSFLM and scaled using SCALA. Resolution in Å.

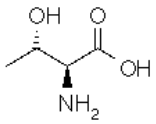
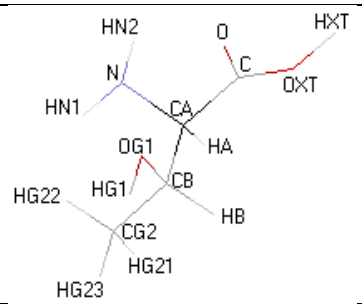
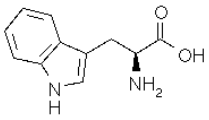
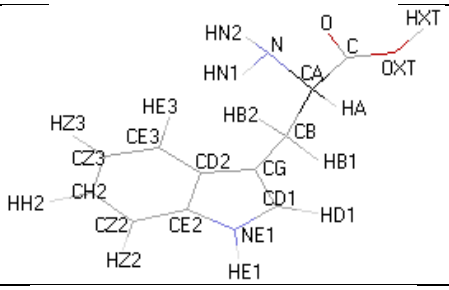
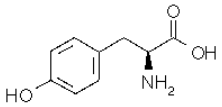
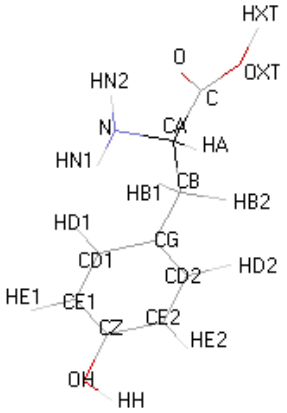
CCS12 A6 **TEST**	5-15% MPD buffer containing 30mM *E. coli* F583	-	-		
CCS13 C3 **TEST**	5-15% MPD buffer containing *15mM E. coli* F583	-	-		
CCS13 C6 **TEST**	5-15% MPD buffer containing 10mM *E. coli* F583	-	-		
CCS12 D5 1 CCS12 D5 2 CCS12 D5 3 CCS12 D5 4 CCS12 D5 5 CCS12 D5 6	5-20% MPD buffer containing 10mM *E. coli* F583. Exchanged immediately. 10/10/12	3m 4m 7m 9m 11m 14m	DC 2.4-3 DC 3.18 - - DC 3.0 -	14/10/12 I02	Unprocessed Fast 3.88 - - Unprocessed -
CCS12 A3 1 CCS12 A3 2 CCS12 A3 3	5-20% MPD buffer containing 10mM *E. coli* F583. Exchanged immediately. 10/10/12	3m 5m 8m	DC 1.4 - DC 2-3	14/10/12 I02	Unprocessed - Fast 2.03
CCS13 D2 **TEST**	5-15% MPD buffer containing 22mM *S. minnesota* R7 13/03/13	-	-		
CCS2 C6 1 CCS2 C6 2 CCS2 C6 3 CCS2 C6 4 CCS2 C6 5 CCS2 C6 6 CCS2 C6 7 CCS2 C6 8 CCS2 C6 9	5-20% MPD buffer containing 22mM *S. minnesota* R7. Exchanged immediately. 15/3/13	1m 3m 4m 13m 15m 18m 21m 24m 26m	No diffraction	20/3/13 I04	-
CCS16 A5 1	5-20% MPD buffer containing 22mM *S. minnesota* R7. Exchanged immediately. 15/3/13	6m	No diffraction	20/3/13 I04	-
CCS16 A1 1 CCS16 A1 2 CCS16 A1 3	5-15% MPD buffer containing 22mM *S. minnesota* R7. Crystals degrading – exchanged before 20% aliquots	1m 2m 5m	DC 1.63 DC 2.13 -	19/5/13 I03	Full 1.75 Fast 2.13 -
CCS16 A1 4 CCS16 A1 5 CCS16 A1 6 CCS16 A1 7 CCS16 A1 8 CCS16 A1 9	As above. Xtals okay. 1µl 20% added. 20µl removed from drop. Total volume of remaining 15% cryo (plus 1µl 100mM $CaCl_2$) added	27m 29m 1h7m 1h10m 1h35m 1h38m	- DC 2.4 - - DC 1.7 DC 2.0	19/5/13 I03	- Fast 2.40 - - Unprocessed Unprocessed

Table 19 Amino acid structures along with their abbreviations, molecular formulae and atom nomenclature

Amino acid	Abbreviation	Molecular formula	Atom nomenclature
Alanine	Ala, A		
Arginine	Arg, R		
Asparagine	Asn, N		
Aspartic acid	Asp, D		
Cysteine	Cys, C		

Glutamine	Gln, Q		
Glutamic acid	Glu, E		
Glycine	Gly, G		
Histidine	His, H		
Isoleucine	Ile, I		
Leucine	Leu, L		

Lysine	Lys, K		
Methionine	Met, M		
Phenylalanine	Phe, F		
Proline	Pro, P		
Serine	Ser, S		

Threonine	Thr, T		
Tryptophan	Trp, W		
Tyrosine	Tyr, Y		
Valine	Val, V